



University of Pennsylvania
ScholarlyCommons

Technical Reports (CIS)

Department of Computer & Information Science

June 1993

A Model of Redundant Information in Dialogue: The Role of Resource Bounds (Dissertation Proposal)

Marilyn A. Walker
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/cis_reports

Recommended Citation

Marilyn A. Walker, "A Model of Redundant Information in Dialogue: The Role of Resource Bounds (Dissertation Proposal)", . June 1993.

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-93-64.

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/cis_reports/278
For more information, please contact repository@pobox.upenn.edu.

A Model of Redundant Information in Dialogue: The Role of Resource Bounds (Dissertation Proposal)

Abstract

This document is a proposal of research intended to complete a Ph.D. in Computer Science. The overall goal of the proposed work is to demonstrate a connection between agents as limited reasoners and the use of informationally redundant utterances in problem-solving dialogues. This document describes some long range objectives and some preliminary results toward this goal. Comments from readers on the proposed work would be most welcome.

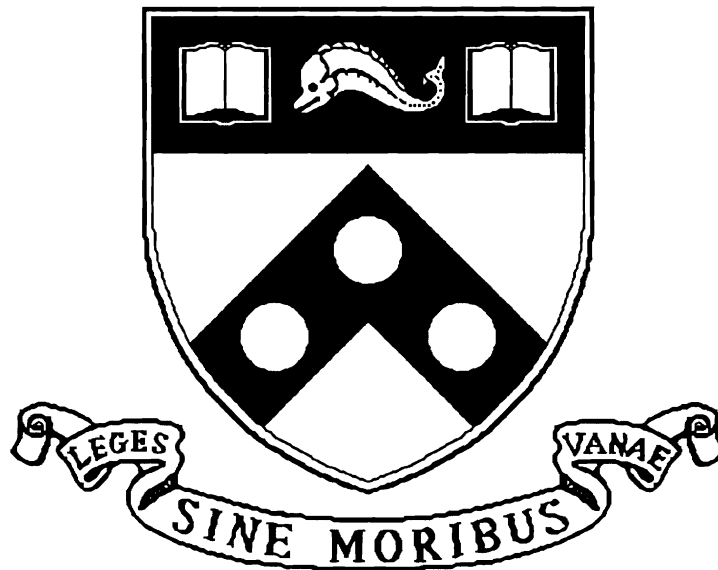
Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-93-64.

**A Model of Redundant Information in Dialogue:
The Role of Resource Bounds
(Dissertation Proposal)**

**MS-CIS-93-64
LINC LAB 253**

Marilyn A. Walker



**University of Pennsylvania
School of Engineering and Applied Science
Computer and Information Science Department
Philadelphia, PA 19104-6389**

June 1993

A Model of Redundant Information in Dialogue:
the Role of Resource Bounds
– Dissertation Proposal –

Marilyn A. Walker¹
University of Pennsylvania
Computer and Information Science
200 S. 33rd St.
Philadelphia, PA 19104
lyn@linc.cis.upenn.edu

¹This research was partially funded by ARO grants DAAG29-84-K-0061 and DAAL03-89-C0031PRI, DARPA grants N00014-85-K0018 and N00014-90-J-1863, NSF grants MCS-82-19196 and IRI 90-16592 and Fellowship INT-9110856 for the 1991 Summer Science and Engineering Institute in Japan, and Ben Franklin Grant 91S.3078C-1 at the University of Pennsylvania, and by Hewlett-Packard Laboratories.

This document is a proposal of research intended to complete a Ph.D. in Computer Science. The overall goal of the proposed work is to demonstrate a connection between agents as limited reasoners and the use of informationally redundant utterances in problem-solving dialogues. This document describes some long range objectives and some preliminary results toward this goal. Comments from readers on the proposed work would be most welcome.

Contents

1	The Paradox of Redundancy	7
1.1	Introduction	7
1.2	Dialogue Type, Autonomy and the Planning Problem	8
1.3	Hypotheses: the initial justification	11
1.4	Overview of the proposal	13
2	Informational Redundancy: Definition and Description	15
2.1	Definition of Redundancy	15
2.2	Descriptive Properties: Intonation and Distribution	19
2.2.1	Intonational Properties	19
2.2.2	Distributional Properties	20
2.3	Entailments	21
2.3.1	Repetitions	21
2.3.2	Paraphrases	23
2.3.3	Modus Ponens Inferences	24
2.3.4	Mathematical Calculations	25
2.4	Presuppositions	26
2.5	Scalar Implicatures	27
2.6	Summary	28
3	Belief Model	29
3.1	Shared Environment Model of Mutual Beliefs	29
3.1.1	Endorsements	31
3.1.2	Combining Endorsements	32

3.2	Saliency of Beliefs	33
3.3	Types of Propositions	33
3.4	Summary	34
4	Attitude	35
4.1	Mutual Belief of Understanding	36
4.1.1	Example of a Repetition	38
4.1.2	Example of a Paraphrase	38
4.1.3	Making Entailments Explicit	39
4.1.4	Summary	40
4.2	Acceptance and Rejection	42
4.2.1	Implicit Rejection and Given/New	44
5	Consequence	47
5.1	Inferring Propositional Relations	48
5.2	Alternates Relation	52
5.3	Summary	54
6	Attention	55
7	Corpus Analysis	59
7.1	Distributional Properties	59
7.2	Claims that are difficult to demonstrate with a distributional analysis	62
8	Computational Simulation	63
8.1	Domain	64
8.2	IRMA architecture for Resource Bounded Reasoning	68
8.3	Belief representation system	71
8.4	Discourse Actions	71
8.5	Agent Communication Strategies	75
8.5.1	Attitude Strategies	78
8.5.1.1	Prediction	79
8.5.2	Consequence Strategies	80
8.5.2.1	Prediction	81

8.5.3	Attention Strategies	81
8.5.3.1	Prediction	86
8.5.4	Summary	86
8.6	Comparison with Other Work	86
8.7	Summary	90
8.8	Acknowledgments	91

Chapter 1

The Paradox of Redundancy

1.1 Introduction

It seems a perfectly valid rule of conversation not to tell people what they already know. Indeed, Grice's QUANTITY maxim: *Do not make your contribution more informative than is required* has often been interpreted this way (Grice, 1967). Stalnaker, as well, suggests that *to assert something that is already presupposed is to attempt to do something that is already done* (Stalnaker, 1978). Thus, the notion of what is *informative* is judged against a background of what is *presupposed*, i.e., propositions that all conversants assume are mutually believed (Lewis, 1969; Schiffer, 1972).

Various formulations of this 'no redundancy' rule are a tenet of many theories of dialogue, from linguistic (Gazdar, 1979), philosophical (Stalnaker, 1978) and computational perspectives (Cohen, 1978; Allen, 1979). However consider the following excerpt from the middle of an advisory dialogue between Harry (h), a talk show host, and Dave (d), his caller:

- (1) (18) h. I see. Are there any other children beside your wife?
- (19) d. No
- (20) h. Your wife is an only child
- (21) d. Right. And uh wants to give her some security

In example 1, the proposition expressed by (20) should already be mutually believed since Harry's assertion simply paraphrases what was said in (18) and (19). It seems unlikely that the truth of (20) is in question, for instead of (19), Dave could not say *No, but she's not an only child*. In addition, (20) is produced with a falling intonational contour typical of declarative statements, rather than a rising one often used to indicate lack of hearing or surprise (Cruttenden, 1986).

Similarly, it would seem that the proposition *The income from the certificate of deposit rules her out as a dependent* should be mutually believed after utterance (7) in the excerpt of a call from Ray (r) below:

- (2) (6) r. uh 2 tax questions. one: since April of 81 we have had an 85 year old mother living with us. Her only income has been social security plus uh approximately three thousand dollars from a certificate of deposit and I wonder what's the situation in terms of claiming her as a dependent, or does that income from the certificate of deposit rule her out as a dependent?
 (7) h. Yes it does
 (8) r. It does
 (9) h. Yup that knocks her out.....

So why does Ray in (8)¹ repeat Harry's assertion of *it does*, and why does Harry paraphrase himself and Ray in (9)? The restatements in 1-(20), and in 2-(8) and (9) above, seem *prima facie* incompatible with efficiency in dialogue and in action in general.

According to common information theoretic accounts, these utterances are INFORMATIONALLY REDUNDANT. I will argue however that they are not communicatively redundant. The definition used to classify them as informationally redundant follows(Hirschberg, 1985)²:

An utterance u_i is INFORMATIONALLY REDUNDANT in a discourse situation \mathcal{S}

1. if u_i has already been said in \mathcal{S}
2. if u_i expresses a proposition p_i , and another utterance u_j that entails p_i has already been said in \mathcal{S}

This definition and the class of data that it covers will be discussed and refined in chapter 2. Here I wish to characterize my approach to determining the function of informationally redundant utterances (henceforth IRUs).

My main interest in the paradox of IRUs is what it shows about the incremental process of two agents constructing the mutual beliefs and intentions that are necessary in order to agree on a plan for future action. IRUs are a useful data set with which to examine the communicative **process** of achieving these mutual beliefs: this is because in the normal case, an utterance adds new information to the context, and so the other communicative functions of the utterance seem to be secondary. However with IRUs no information is added. This means that the other functions that these utterances achieve in dialogue are more apparent.

1.2 Dialogue Type, Autonomy and the Planning Problem

This section characterizes the type of dialogues that the analysis is based on and the goals that agents engaged in these dialogues are trying to achieve. I also discuss properties of these agents that contribute to their success or lack of success in achieving these goals, as well as the methods by which they go about constructing plans to achieve their goals.

¹Utterance (8) is also said with a falling intonation contour typical of a declarative statement.

²See also Hobbs' definition of elaboration(Hobbs, 1979).

The initial analysis is based on a corpus of naturally occurring advisory dialogues from a talk show for financial advice, *Speaking of Your Money*³. I will refer to this as the HG (Harry Gross) corpus. IRUs do occur in single speaker discourse, and their function may be much the same as those that will be claimed here. However there are two reasons to focus on dialogue: (1) IRUs seem to be much more prevalent in dialogues, and (2) IRUs seem to be more prevalent in problem solving situations.

In the talk show setting, I assume the agents have a mutual goal. The mutual goal is to find a satisfactory plan for future action that addresses the problem or query presented by the caller.⁴ This goal is normally stated explicitly in the first full turn of the caller. Furthermore, the total set of goals in the corpus are restricted: they always contribute to one of a small number of higher level goals. The assumed general goal is to *maximize assets* such as income, and the caller's opening statement limits the way in which s/he is currently considering achieving this goal, such as by *maximizing return on investments* or by *minimizing taxes*.

This means that the planning situation is similar to the one characterized in Pollack's treatment of plan inference, namely PLAN EVALUATION(Pollack, 1986; Pollack, 1991). Pollack focused on the inference of domain plans, in situations where the goal has been made explicit. The reasoning in the situations she discusses involves the evaluation of a domain plan, when it is not assumed that agents have identical beliefs about the domain. This contrasts with other work on plan inference in which the goal of the inference process is to determine the domain plan of the acting agent(Kautz, 1990).

However, in the plan evaluation case, even though there is no need to infer the domain goal, agents must reason from another conversant's utterances, to beliefs about their mental state. In an extended plan evaluation dialogue, utterance level intentions must be inferred, as well as the role that these utterances play in achieving the mutual goal(Grosz and Sidner, 1986; Litman and Allen, 1990).

Agents must achieve mutual belief that they have constructed a plan to address the mutual goal. According to the beliefs model proposed in chapter 3, mutual beliefs are publicly constructed by two agents engaged in dialogue, but based on a number of underlying assumptions more or less endorsed. This defines a weak model of mutual beliefs; mutual beliefs may be as weak as a mutually salient and public supposition. In the joint problem solving task examined here, what must be achieved is that there is mutual belief as to whether an utterance and its propositional content have been:

- Understood as the speaker intended
- Accepted or Rejected
- Incorporated into the evolving plan to address the mutual goal

None of these mutual beliefs can be automatically assumed. They must be achieved by agents who are both resource-bounded and autonomous, with their own prior beliefs, preferences and goals.

³This talk show is on WCAU in Philadelphia. The corpus was taped from a radio broadcast and originally transcribed by Pollack and Hirschberg(Pollack et al., 1982; Walker and Whittaker, 1990).

⁴The agents may have additional goals as well such as being entertaining, being clever, flirting, etc.

Agents' limitations and the fact that they do not share all the same domain beliefs mean that understanding the utterance and determining how its content plays a role in the evolving plan cannot be assumed. Agents' autonomy means that the acceptance of beliefs and candidate plans also cannot be assumed. Acceptance is distinct from just understanding what has been conveyed, as I explain further in chapter 4.⁵ Autonomous agents must decide whether they will accept as a belief some fact that they are told and whether they wish to accept as a goal something that is suggested to them. Agents may make these decisions by reasoning about different ways to achieve the mutual goal. And since the goal is mutual, agents also must resolve conflicts in beliefs and intentions (Galliers, 1989).

In sum, the achievement of these mutual beliefs is not guaranteed. This is in contrast to certain accounts of speech act interpretation (Litman and Allen, 1990; Cohen and Levesque, 1991) where acceptance follows directly from the speaker's utterance. I will use the term *COLLABORATIVE PLAN* or 'plan to achieve the mutual goal' to describe the constellation of mutual beliefs and intentions that agents who are engaged in a joint problem solving task must achieve (Whittaker and Stenton, 1988; Walker and Whittaker, 1990).⁶

Rosenschein has shown that planning between two agents is guaranteed to converge on a collaborative plan only when the information shared between the agents increases monotonically (Rosenschein, 1985). In the nonmonotonic case, even when there are no inter-agent conflicts, it isn't possible generally to guarantee that the agents will agree on a plan. However there is one discourse situation where plan convergence can be guaranteed in the nonmonotonic case. This is when one agent's knowledge is a superset of the other agent's knowledge, as in Grosz's pump assembly task or the task in the trains domain (Grosz, 1977; Traum, 1991). Yet this situation is not the typical one for most problem solving dialogues in real life (Kidd, 1985; Whittaker and Stenton, 1988; Walker and Whittaker, 1990).

The dialogue situation examined here is an instance of the nonmonotonic case; agents in dialogue reason as to whether to accept new incoming information or reject it, and whether, as a result, to disbelieve something they previously believed (Galliers, 1991a). Yet despite this nonmonotonicity, much of the time the conversants do manage to agree on a plan to achieve the mutual goal. The question is how they do it and what IRUs show about this process.

This section has introduced the paradox of informationally redundant utterances (IRUs) and described the perspective from which I will address this paradox. I emphasize that these utterances are not communicatively redundant, even though they are informationally redundant. Section 1.3 will describe my hypotheses about the roles that IRUs play in dialogue. These hypotheses assume that agents have certain cognitive properties. Section 1.3 examines these cognitive properties, the way that the hypotheses follow from them, and the way that I intend

⁵However, understanding may be a prerequisite for acceptance, and in certain tasks, such as achieving reference (Clark and Wilkes-Gibbs, 1986), there is no need for such a distinction. This is probably why some researchers use the term acceptance for the achievement of understanding (Clark and Schaefer, 1989; Brennan, 1990).

⁶This term has much in common with Power's characterization of mutual intention, with Grosz and Sidner's notion of a *SharedPlan*, and with Cohen and Levesque's concept of *Joint Intention* (Power, 1979; Power, 1984; Grosz and Sidner, 1990; Cohen and Levesque, 1991). For the time being, it will be simpler if I use my own terms. The approach that I outline here extends, rather than conflicts, with these other approaches. Section 8.6 will discuss these accounts further.

to support the hypotheses.

1.3 Hypotheses: the initial justification

IRUs leave us with a paradox given the fact that agents do not generally perform actions that have no purpose. Assuming that IRUs actually must have a function in discourse, the question is what communicative functions do IRUs achieve?

This section describes hypotheses about three communicative functions that agents can achieve through IRUs. These functions demonstrate that the conceptualization of dialogue as a simple process of information exchange is too limited. I will discuss some cognitive properties of human agents and argue that my hypotheses are motivated by these properties. I will also discuss how I will attempt to validate these hypotheses.

I propose that IRUs have three general functions:

- Attitude: to augment the evidence supporting beliefs about mutual understanding and acceptance
- Consequence: to augment the evidence supporting mutual beliefs that certain inferences are licensed, or to provide a context in which an inference is more easily made
- Attention: to manipulate the locus of attention of the discourse participants by making or keeping a proposition salient

Individual IRUs may simultaneously address one or more of these functions. For instance, every utterance, redundant or not, will claim the locus of attention at the point when it is said. Specific examples of these three functions will be discussed later. Here I wish to consider why it is that agents might want to achieve them.

These hypotheses partly follow from some well attested demonstrations about the limitations of agent's processing capabilities. When we consider how agents are resource-bounded, it is clear that human agents have limited processing capacity(Kahneman et al., 1982). They must often respond in dialogue in a timely manner, and yet they need time to search memory, make inferences and decide what to say. They also have limited attentional capacity(Miller, 1956; Anderson and Bower, 1973). Therefore an agent can't always keep all the relevant facts in mind, and also can't make all the inferences that follow from these facts.

Agent's resource-bounds show up in the way they behave. Since dialogue is an instance of action in general, there should be phenomena in dialogue that are a result of the fact that agents are resource-bounded. First, agents must manage their own resources so that dialogue proceeds smoothly. Second, in extended plan evaluation dialogues, agents may have to reason about other agents' limitations. Because agents depend on other agents to realize the mutual goal, the limitations of one agent can affect the other. This means that in her own self-interest, an agent may actually model the resource limitations of another and act on this model, in order to decrease the amount of processing the other agent must do and thereby ensure that the conversation goes smoothly(Clark and Wilkes-Gibbs, 1986).

My hypothesis is that IRUs are a surface manifestation of agents' management of their resource limitations.

The need to achieve goals related to Attitude follows from the fact that agents are autonomous with their own preferences, beliefs, and goals. Therefore an agent will not necessarily believe what she is told or adopt a goal that is suggested to her. Furthermore, in contrast to agent's physical activities, which are observable, the mental objects of other agent's beliefs, preferences and intentions are non-transparent. Therefore agents look for and give public evidence of the effects of utterance actions on their beliefs and intentions. Chapter 4 will discuss Attitude in more detail.

The need to achieve goals related to Consequence follows from the fact that agents are not logically omniscient. They might not have time to make all the relevant inferences and might not know all the relevant inference rules (Konolige, 1985). In addition, there are always potentially more inferences that can be made at a particular point in a dialogue. Thus agents may make relevant inferences explicit. This can indicate to other agents that a certain inference was made, as well as ensuring that another agent made a desired inference. Chapter 5 will discuss Consequence in more detail.

Since human agents have limited attention, the need to manipulate Attention also follows from agents' resource-bounds (Grosz, 1977; Sidner, 1979; Clark and Clark, 1977). Agents need to coordinate in dialogue and one aspect of this coordination is tracking and manipulating the locus of attention of other agents. At times, utterances that are informationally redundant serve to set the context of the discourse and ensure that the agents involved are jointly attending to the same concepts (Whittaker and Stenton, 1988; Walker and Whittaker, 1990). Chapter 6 will discuss Attention in more detail.

Although I have separated my hypotheses into separate Consequence and Attention functions, there is an intimate relationship between processing time and attention; items currently attended to can be accessed with constant or zero time (Anderson, 1974; Garrod and Trabasso, 1973), whereas search is required for other items. If inferences are focused on what is currently attended to, as seems likely (Kahneman et al., 1982), then limits on inferencing follow directly from limits on attention. To avoid the problem of search for items not attended to, agents may manipulate the context to direct inferences to be about a particular topic. The data also support the view that certain inferences seem to depend on two or more facts being attended to at the same time (see section 5.1). If an agent says something to make sure that a fact is currently attended to, because a desired inference is dependent on that fact, that utterance functions on both the Attention and the Consequence dimensions.

I will not discuss these hypotheses further until Chapter 4. The important point to keep in mind is that goals to achieve these functions follow from agents' autonomy and resource limitations. This viewpoint will come up repeatedly in what follows. In order to provide evidence for these hypotheses I will discuss distributional properties of the data in chapter 7. However for various reasons discussed there, the distributional analysis only provides weak support for the hypotheses. Chapter 8 describes the design of a computational experiment that is intended to provide further support.

1.4 Overview of the proposal

Section 1.1 has introduced the paradox of informationally redundant utterances (IRUs) and described the perspective from which I will address this paradox. Section 1.3 described my hypotheses about the roles that IRUs play in dialogue and the associated cognitive properties.

Because the idea that agents actually produce IRUs seems to go against many previous accounts of how language works, it is important to define very carefully what counts as an IRU. Therefore I devote chapter 2 to a precise definition of an IRU. This definition is based on different logical properties of propositions in the discourse situation. Section 2.2.2 discusses specific examples that instantiate the definition given in section 2.1. These examples are organized by the logical properties of IRUs rather than by the hypothesized communicative functions. The purpose of this section is to familiarize the reader with the range of the phenomena. Section 2.2 also introduces a number of distributional parameters that are used to classify the examples; these parameters are used in chapter 7, which discusses an initial distributional analysis of the data in the corpus.

Chapter 3 proposes a model of mutual beliefs for dialogue that is both motivated by and motivates the data. There are three core distinctions about the status of a belief that are supported by this model: (1) whether or not the belief is salient, (2) what type of evidence endorses the belief, (3) whether or not this endorsement has been made public in the dialogue. Once the phenomena of redundancy and the model of mutual beliefs have been described, the following chapters return to the discussion of the hypotheses.

Chapter 4 elaborates on the Attitude hypothesis given in section 1.3. This chapter proposes an inferential account of how understanding and acceptance are achieved that is based on the model of mutual beliefs presented in chapter 3. Chapter 4 makes the distinction between understanding and acceptance and discusses cases of IRUs that support the inference of mutual beliefs about understanding and acceptance. A model of this inference process is proposed and the class of examples explained by this model is demonstrated.

Chapter 5 on Consequence and chapter 6 on Attention give motivating examples of IRUs that are hypothesized to demonstrate these functions. The Consequence function depends partly on an account of a resource-bounded agent based on the IRMA architecture (Bratman et al., 1988; Pollack and Ringuette, 1990). The Attention function will be developed along the lines of Grosz and Sidner's work on attentional state, and the concept of a focus space (Grosz, 1977; Sidner, 1979; Grosz and Sidner, 1985).

Chapters 7 and 8 discuss what will be done to validate the hypotheses that have been proposed. Chapter 7 describes an initial corpus analysis using the parameters introduced in section 2.2.2, and discusses what can be tested with this methodology. Chapter 8 proposes a computational simulation of a dialogue between two agents that can provide a testbed for the proposed hypotheses.

Chapter 2

Informational Redundancy: Definition and Description

This chapter discusses in more detail what utterances in a dialogue count as IRUs. The main purpose of this definition is to delimit a data set that this thesis will focus on.

Typically in language, speakers are more or less explicit about what exactly they take the current context to be and what exactly they mean. Levels of explicitness range from utterances consisting of all new information to utterances that consist of all old information. On this continuum, I focus on clauses that did stand alone, or could stand alone, as utterances. Some of these clauses are part of a fuller turn.

In addition, I only focus on the informational component of an utterance, the propositional content. Utterances consist of a string that is said, the proposition that is realized by that string in a particular context, and an associated utterance level intention, which consists of the role of that utterance in the overall structure of the discourse and the conversants' intentions. The notion of IRU is only meant to refer to the propositional content of an utterance. These utterances are not communicatively redundant, and an IRU does not necessarily realize the same utterance level intention as the utterance that originally added the propositional content of the IRU to the discourse situation.

2.1 Definition of Redundancy

This section provides a definition of redundancy that will be used throughout the remainder of the proposal. Section 2.2.1 provides some of the intonational parameters that will be used to describe the data given in the following sections.

The definition of informational redundancy is based on different logical properties of propositions in the discourse situation. This definition of informationally redundant depends on a “straw man” view of language. It is based on a view of communication in which (1) an agent merely saying an utterance adds the propositional content of that utterance to mutual beliefs, (2) one proposition is communicated a time, and (3) all the inferences deriving from this

proposition in combination with all the previously communicated propositions are automatically derived (Gazdar, 1979; Barwise, 1988). The discerning reader will note that this account contravenes several of the agent limitations and properties described above. My argument will be that the function of IRUs can only be derived in a model that recognizes these agent properties and limitations. For convenience, I repeat here the definition of redundancy given in section 1.1:

Definition of Redundancy: Version 1

An utterance u_i is INFORMATIONALLY REDUNDANT in a discourse situation \mathcal{S}

1. if u_i has already been said in \mathcal{S}
2. if u_i expresses a proposition p_i , and another utterance u_j that entails p_i has already been said in \mathcal{S}

An utterance is defined as a clause, or a phrase in cases when there is no finite verb in an utterance, such as in elliptical noun phrases. Condition (1) of the definition means that saying an utterance in a discourse situation adds the propositional content of that utterance to the discourse situation. The second condition depends on being able to identify what is entailed from what is said; it relies on concepts such as presupposition, paraphrase, and logical inference.¹

It will be useful to have a term to refer to the utterance(s) that originally added the propositional content of the IRU to the discourse situation. Following work on referential discourse entities, I will call this the IRUs ANTECEDENT. Actually I will use the term antecedent to refer to both the prior utterance and the proposition realized by that prior utterance, but this should not cause any confusion. In the dialogue excerpts given here, IRUs will be marked with CAPS whereas their antecedents will be given in *italics*. An IRU may be explicitly related to its antecedent, e.g. a repetition, or implicitly related to its antecedent, e.g. inferable by modus ponens. The types of IRUs examined here are shown in figure 2.1.

TYPE of IRU	EXPLICIT Relation to Antecedent(s)	IMPLICIT Relation to Antecedent(s)
Entailment	Repetitions Paraphrases	Modus Ponens Inferences Mathematical calculations
Presupposition		Existential Presuppositions
Conversational Implicature		Scalar Implicatures

Figure 2.1: Types of Informationally Redundant Utterances

¹For the time being I don't include information in the discourse situation that could be jointly perceived by two agents in a shared **physical** environment, such as *Your knee is touching mine*.

Thus there are three basic types of IRUs: entailments, presuppositions and implicatures. I will discuss their logical properties and the motivation for including them as IRUs in the remainder of this section.

In the initial corpus analysis, what counts as an entailment is very strictly defined. I assume that lexical knowledge is shared if it is not domain specific. Otherwise all the information that an entailment depends on must be made public in the dialogue, although I will assume knowledge of inference rule schemas such as *modus ponens*. While it is possible that some background knowledge is shared in this dialogue situation and while generally conversants do share various kinds of background knowledge, if this knowledge has not been made public in the dialogue, for this analysis I do not assume it to be shared, because it isn't possible to know for sure that it is. As mentioned before, the purpose of this definition is to select a set of data to focus on, and it is important for the selection criteria to be replicable.

Strictly speaking, neither presuppositions nor implicatures are typically counted as entailments of utterances since they are not part of the truth conditional meaning of an utterance. Thus they do not fit under the definition of redundancy given above. According to the classical Gricean view, there are two logical properties that distinguish presuppositions and implicatures from entailments (Levinson, 1983; Sadock, 1978):

- Reinforceability: whether the inference can be made explicit without redundancy
- Defeasibility: whether the inference can be defeated by additional information or depending on the discourse situation when the utterance is made

These properties have ramifications for the beliefs model for dialogue (Sadock, 1978; Grice, 1975). The way in which the types of IRUs given in Figure 2.1 are classified by these properties is shown in Figure 2.2 (Bridge, 1991).

	Reinforceability	Defeasibility
Entailment	no	no
Presupposition	order dependent	yes
Conversational Implicature	yes	yes

Figure 2.2: Properties of different types of Information Antecedents

In the next section, I will present a number of examples that one would not expect to occur given these logical properties, that is counter-examples to the classical Gricean view. However I will briefly discuss here the basis of this classification. For example, contrast 3a and 3b with 4a and 4b.

- (3) a. # My sister is older than I am and I am younger than my sister.
 b. # My sister is older than I am but I'm not younger than my sister.

Example 3a conjoins two clauses where the second is entailed by the first. Example 3b conjoins two clauses where the second one attempts to defeat the first. Because entailments are neither reinforceable nor defeasible, examples 3a and 3b are anomalous. The anomaly occurs independent of the order in which these clauses are stated.

Like entailments, presuppositions are not supposed to be reinforceable (Gazdar, 1979). Presuppositions are similar to implicatures in that they are not part of the truth-conditional meaning of an utterance, but unlike implicatures they are not defeasible. Existential presuppositions are often carried by the use of definite expressions. For example, an utterance in which I speak of *my sister* will carry an existential presupposition that *I have a sister*. Existential presuppositions also survive negation.

- (4) a. I have a sister and my sister is older than I am.
 b. # My sister is older than I am and I have a sister.
 c. # My sister is older than I am but I don't have a sister.

In example 4a, the presupposition is in the clause before the one that presupposes it. Since presuppositions can be reinforced as long as the reinforcement comes before the clause that introduces the presupposition, 4a is not anomalous. However 4b puts the presupposition after the clause that presupposes it and is anomalous. Example 4c shows that presuppositions are not defeasible.

Conversational implicatures, such as the inference from *some* to *not all* are both reinforceable without anomalous redundancy as in 5a, as well as defeasible as in 5b:

- (5) a. I ate some of the cookies but I didn't eat all of them.
 b. I ate some of the cookies and in fact I ate all of them.

Since conversational implicatures are reinforceable it may seem odd to include as IRUs utterances that have a conversational implicature as an antecedent. A natural account would be that the IRU is just a reinforcement of an implicature. However all the work that I am familiar with on the properties of implicatures, such as reinforceability and defeasibility, has been based on **single utterances** such as those in example 5, where the defeater or reinforcer is said immediately by the same speaker. Since much of the data examined here provides counter-examples to the classical view, I include reinforcements of implicatures when said by other speakers, or when non-adjacent to their antecedent, in the class of IRUs examined here. I keep the restriction of IRUs non-adjacent to their antecedent to leave aside the cases that have been studied previously as defeating or reinforcing an implicature. Implicatures as antecedents will be kept in separate tallies so there should be no confusion.

There is additionally another class of utterances that seems as though they should be classified as informationally redundant. These are PROMPTS such as *uh huh*. Prompts provide no new propositional content to the dialogue, and thus are a paradox in the same way that IRUs are. However prompts don't have an antecedent in the dialogue; no matter what has been

said previously, a prompt can add no new information. Because prompts share a number of properties with IRUs as to their function in dialogue (Schegloff, 1982; Whittaker and Stenton, 1988; Walker and Whittaker, 1990), and because they add no new information, I will include them as a type of IRU in the remainder of the proposal.

The previous discussion of presuppositions, implicatures and prompts motivates a revision of the definition of redundancy that will be used in the remainder of the proposal.

Definition of Redundancy: Final Version

An utterance u_i is INFORMATIONALLY REDUNDANT in a discourse situation \mathcal{S}

1. if u_i has already been said in \mathcal{S} ,
2. if u_i expresses a proposition p_i , and another utterance u_j that entails p_i has already been said in \mathcal{S} ,
3. if u_i expresses a proposition p_i , and another utterance u_j that presupposes or implicates p_i has already been said in \mathcal{S} either non-adjacent to u_i or by another speaker,
4. if u_i has no propositional content, (prompts such as *uh huh*).

This is a broad definition of IRUs, but the motivation for collapsing these types together is to simplify the discussion of utterances in which the information provided by the utterance could reasonably be expected to already be ‘available’ in the conversation. Each of these categories will be tracked separately so that there is no potential for confusion as to the claims being made.

This section defines what counts as an IRU for the purpose of selecting a particular set of data that this thesis focuses on. Specific examples will be given in section 2.2.2 and the way in which different types of examples relate to the hypothesized functions discussed in section 1.3 will be explored in chapters 4, 5, 6, 7 and 8. However, before discussing the examples, I will briefly digress to explain some terms that that will be used to describe intonational properties of the corpus.

2.2 Descriptive Properties: Intonation and Distribution

2.2.1 Intonational Properties

The initial data analysis includes the examination of a number of intonational properties of IRUs. I examine the intonation with which an IRU is produced because it is obvious that the same string of words can be said in many different ways, and thus communicate many different meanings (Lieberman and Sag, 1974). I will not use intonation as a defining feature of IRUs, relying on the logical properties given in the previous section. However the intonation with which an IRU is produced will be relevant when it comes to discussing whether a particular IRU has a particular communicative function.

One way that intonation is relevant is in distinguishing between an utterance with a declarative syntactic form that is said with a rising ‘question’ intonation in comparison with the same string said with a falling ‘declarative’ intonation. This distinguishes cases in which the IRU could be seen to function as a clarifying question or to indicate lack of hearing (Cruttenden, 1986).

The second way that intonation is relevant follows from the classical view that each utterance can be divided into old information and new information (Halliday and Hasan, 1976; Clark and Haviland, 1977; Prince, 1981b). Old information is already in the discourse situation, what I might call redundant information. However the classical view is that each utterance has at least one item of new information, the information focus. Since the IRUs examined here provide no new information, one aspect of this study is to investigate their potentially anomalous intonational realization, and the ramifications of this for the classical view of information structure.

There are three different terms that I will use to describe intonational properties of utterances:

- **Boundary tone:** this term describes whether the utterance ends with a rising tone, a falling tone or a level tone
- **Broad Focus:** when an utterance is said with broad focus, no particular element of the utterance is selected for special attention, or the whole utterance is so selected.
- **Narrow Focus:** when an item in an utterance receives narrow focus, the utterance is normally interpreted as making a predication about that item. In addition, some types of narrow focus give rise to certain inferences. This means that an IRU, said with narrow focus, may add inferences to the discourse situation.

I will use these terms in the rest of the proposal as part of the description of IRUs.

2.2.2 Distributional Properties

The following sections will discuss instances of IRUs from the HG corpus to illustrate the broad range of examples. These examples will be discussed in terms of a number of distributional parameters. These parameters are:

- the logical relation of the IRU to its antecedent.
- whether the IRU is adjacent to its antecedent, or the antecedent is remote.
- whether the antecedent was said by the same or by another speaker.
- whether the IRU is said with a falling boundary tone as though it were a declarative utterance, or a rising boundary tone.

.

These are not the only parameters of interest, but they provide a broad classification of the data. These parameters will be important in discussing the functions of redundancy in the following sections.

The parameter of the logical relation of the IRU to its antecedent is the basis of the following sections' organization. Section 2.3 discusses IRUs that are related to their antecedent(s) via entailment, both explicit entailments such as repetitions in section 2.3.1, and implicit ones such as mathematical calculations (sections 2.3.3 and 2.3.4). Section 2.4 discusses presupposition reinforcement, and section 2.5 discusses reinforcement of scalar implicatures.

2.3 Entailments

This section describes instances of entailments as IRUs. First I will discuss the entailments that bear an explicit relation to their antecedents, such as repetitions and paraphrases, and then those entailments that have an implicit relation to their antecedents such as various types of inferences.

2.3.1 Repetitions

REPETITIONS are a trivial subset of entailments that match the first condition in the definition of redundancy given in section 2.1. Characteristics of repetitions include the stipulation that the word order is the same in a repetition as in its antecedent. However, I do assume that repetitions include utterances in which a surface mapping has occurred between first and second person deictics across speakers, e.g. *I* can be mapped to *You*. Consider the repetition in (9) below, extracted from a longer dialogue.

- (6) (8) h. you can stop right there: *take your money*
 (9) j. TAKE THE MONEY
 (10)h. absolutely.....

In example 6, utterance (9) is said with a falling intonation pattern typical of declarative utterances. In general, the intonation with which these repetitions are performed seems to be a critical part of their meaning. Stylized falls with phrase final Mid tones are common on adjacent repetitions and paraphrases (Liberman, 1975; Ladd, 1980; Liberman and McLemore, 1992). Rises may indicate that the information provided is being questioned in some way. For example in 7.12, the repetition is said with a rising 'question' contour.² The information that is apparently being 'questioned' in 7.12 has presumably already been provided in 7.11.

- (7) (10) h. How do you feel about stock?
 (11) m. Well, alright, *give it a try*.
 (12) h. GIVE IT A TRY?
 (13) m. Yes.
 (14) h. How about if we go for twenty five thousand dollars in three different southern utilities stocks?

²The function of rises that I assume here may be specific to this corpus (McLemore, 1991).

22CHAPTER 2. INFORMATIONAL REDUNDANCY: DEFINITION AND DESCRIPTION

The previous two examples are cases where a speaker repeats a proposition whose antecedent is **adjacent** to the repetition, i.e. was said by another conversant in the just prior turn. Repetitions across speakers can also occur in cases where the antecedent is **remote** from the IRU, as in (43) and (58) below.

- (8) (42) h. Now what is your income situation?
 (43) m. *We're both retired* and our income for the year is about um 24.. about 26 thousand
 (44) h. Have you other securities than stock? Have you any bonds or certificates?
 (45) m. Yes yes we do- we have some certificates oh about uh 15 - 20 thousand not much we're not rich - and we have a house completely paid for, have some land in the poconos, completely paid for - and uh that's actually the extent of our uh
 (46) h. Ok - on the proceeds of that gm stock
 (47) m. Yes
 (48) h. I'd like to see you put that into two different southern utilities
 .
 (clarification of southern utilities)
 .
 (58) h. And those I think that you will find that that will give you a good return, **YOU ARE RETIRED**
 and that is primarily what you are looking for addition I think that eventually those stocks will rise as well
 (59) m. uh huh

All the examples seen so far are cases where a speaker repeats an utterance that **another** conversant originally added to the context. In 9, the speaker repeats her own statement, which was initially added to the context by the first clause, and then presupposed by the second clause (See also (Horn, 1991; Ward, 1985; Ward, 1990)). This repetition is completed within a single speaker's turn.

- (9) *I like you* Liz.
 I don't know *why I like you*. But I **LIKE YOU**. (CS, 3/4/92)

Adjacent repetitions are frequently elliptical; only a portion of the previous utterance is selected for repetition in 10. The repetition of any subconstituent or any part of what was said counts as a repetition.

- (10) (24) r. uh no. but how about the twenty eight hundred dollars interest?
 (25) h. *the amount of interest you can put into an ira*³
 (26) r. I I CAN and
 (27) h. yes

³An IRA is an Individual Retirement Account, which is a way of putting aside some income and deferring paying income taxes on it.

Finally, the proposition conveyed by the repetition may be embedded within another propositional operator, such as *if* in example 11.

- (11) (30) h. *is he self-employed?*
 (31) l. *yes*
 (32) h. i'm sorry, i missed that
 (33) l. *yes he is.*
 (34) h. if HE'S SELF-EMPLOYED, why not start for him a Keogh plan⁴?
 you can't get an IRA after 70, but you can get a Keogh plan

In 11, the utterance of (34) has an embedded proposition, *he's self-employed* that is already both given and salient in the discourse.

2.3.2 Paraphrases

With PARAPHRASES, as with repetitions, the relation of the paraphrase to its antecedent may be one of a number of types, both logically and sequentially. Some are adjacent to their antecedent and some are remote, some occur across speakers whereas sometimes a speaker paraphrases herself.

The simplest paraphrases are slightly more complex than a repetition, involving syntactic manipulations such as topicalization(Prince, 1981a; McKeown, 1985). These consist of repackaging the same information into a different form. For instance, this type of rule can generate (b) from (a):

- (12) a. You cannot claim the dependency.
 b. The dependency you cannot claim.

These two utterances are different with respect to their discourse function, but their propositional content is the same(Ward, 1985; Prince, 1986; Ward and Prince, 1991; Prince, 1985). Another type of paraphrase is inferences based on lexical or phrasal knowledge(Melčuk, 1988). For example in 13 below:

- (13) a. There are no other children beside your wife.
 b. Your wife is an only child.

The inference from (a) to (b) is based on an axiom that defines *only child* as equivalent to *someone with no siblings*. This was demonstrated in example 1 in section 1.1. Similarly in the first clause in (56) below, the lexical semantics of the verb *separate* entails a division of something into at least two distinct parts. However Harry paraphrases himself in the following utterance.

⁴A Keogh plan is another way of putting aside some income and deferring paying income taxes on it. However Keogh plans are only available to self-employed individuals.

- (14) (56) h. and *but separate it*. I DON'T WANT IT ALL IN ONE.

A slightly more complex set of paraphrases involves logical transformations such as the mapping between (a) and (b).

- (15) a. All of the money you have spent is indirect.
b. None of the money you have spent is direct.

This is demonstrated in the excerpt below in which Harry (h) paraphrases his statement in (11) with the utterances in (13) and (15).

- (16) (9) h. ... Now there is something you can do. Do you support her in any way?
(10) r. yes well i mean she yeah we supply, you know, everything, heat light uh food.
In other words we, you know, she pays nothing as far as the uh upkeep of the home.
(11) h. and *the only amount that that you have spent then is indirect*.
(12) r. uhh
(13) h. THERE'S NOTHING DIRECT
(14) r. yea
(15) h. THAT YOU SPEND
(16) r. food

2.3.3 Modus Ponens Inferences

Entailments also include inferences made via classic inference rules, such as MODUS PONENS and MODUS TOLLENS. Condition (2) of the definition of redundancy means that if (a) and (b) below are in the discourse situation, then (c) will be an IRU since it is entailed from (a) and (b) via MODUS TOLLENS:

- (17) a. You can buy an I R A if and only if you do NOT have an existing pension plan.
b. You have an existing pension plan.
c. You cannot buy an I R A.

The following excerpt demonstrates this structure. Utterance 18.(15) realizes the proposition in 17a, utterance 18.(16) realizes the proposition in 17b, and utterance 18.(17) makes the inference explicit that is given in 17c for the particular tax year of 1981.

- (18) (15) h. Oh no. *I R A's were available as long as you are not a participant in an existing pension*
(16) j. Oh I see. Well I did work, *I do work for a company that has a pension*
(17) h. ahh. THEN YOU'RE NOT ELIGIBLE FOR EIGHTY ONE

I also assume that there is an entailment scale for quantifiers and amounts (Horn, 1984). For example, 19(a) entails 19(b). So if 19(a) is in the discourse situation, a statement of 19b will be an IRU.

- (19) a. I have 10,000 dollars in a certificate of deposit.
 b. I have some money in a certificate of deposit.

The following excerpt demonstrates making an inference explicit that follows from this type of inference rule. Utterance (28) realizes 19a, and utterance (31) realizes 19b.

- (20) (28) j. *I also have ten thousand in a cd*⁵ six month CD
 (29) h. Oh, uh hang on, do you have anything else?
 (30) j. That that's it. that's it.
 (31) h. Well now we're talking about something slightly different.
 YOU ALREADY HAVE SOME MONEY IN A CD.
 Have you anyone dependent on you?

2.3.4 Mathematical Calculations

I will also assume that entailments include some simple mathematical calculations that would lead for instance from (a) and (b) to (c), assuming some appropriate definition of *about*.

- (21) a. I am now making 9 thousand 7 hundred dollars.
 b. I used to make about 17 thousand dollars.
 c. I am now making about half of what I used to make.

This is demonstrated in (4), (13) and (14) below.

- (22) (4) c. oh sir - i was put off the job, i had a city job, and uh they put me off with high blood pressure and hypertension. *so i get a a pension retirement of 9 thousand 7 hundred dollars*, that's non-service connected. i have to pay income tax on that don't i?
 (5) h. fred?
 (6) c. i have to pay income tax ...
 (7) h. yea yes
 .
 .
 (12) f. well uh - how does that amount relate to the amount you were receiving when you were employed

⁵A CD is a Certificate of Deposit, where money is put aside for a period of time at a higher rate of interest.

- (13) c. *uh i was making approximately about 17 thousand*
- (14) f. ok - AND YOU'RE NOW GETTING ABOUT 50 % OF THAT FIGURE
- (15) c. uh yes sir.....

This concludes the description of IRUs related to their antecedents via entailment. The next two sections discuss IRUs that are related to their antecedents through the non-truth-conditional inference types of presupposition and implicature.

2.4 Presuppositions

As discussed in section 2.1, like entailments, presuppositions are not supposed to be reinforceable. Presuppositions are similar to implicatures in that they are not considered to be part of the truth-conditional meaning of an utterance; unlike implicatures they are not defeasible. The EXISTENTIAL presupposition given in 23(c), is presupposed by both 23(a) and 23(b).

- (23) a. The charges will be excessive.
- b. The charges won't be excessive.
- c. There are charges.

However consider the following excerpt.

- (24) (22) b. are there ah .. i don't think *the ah brokerage charge* will be ah that excessive
- (23) h. no *They're* not excessive but THERE ARE CHARGES

In 24, it seems that Bill (b) might have been starting out with a question as to whether there are brokerage charges or not. However even if this question were completely expressed, what it would add to the representation of the discourse situation is that *Bill doesn't know whether or not there are brokerage charges*. This is cancelled by the next part of Bill's utterance, which uses a definite referring expression *the brokerage charge*. In (23) Harry (h) further presupposes the existence of said brokerage charges, by predicating of them that they are in fact not excessive, then affirms their existence.

Similarly in (13) below the caller (c) asserts that *there is a family*, predicates of this family that they were not close to *him*, the caller's former employer, and then affirms again that *there is a family*.

- (25) (10) h. let me ask one question of you
- (11) c. yup
- (12) h. uh was there a family
- (13) c. yes, well that that is the problem. *there is a family*. uh it it *they* ((pause))

weren't close to him because he was really a loner. hmmm, but THERE IS A FAMILY INVOLVED to this matter of fact the umm, well i'll explain what happened.

It is not clear on first inspection whether these affirmations actually have to do with affirming the existence of the referent, whether they are related to manipulating attentional state, or they have to do with some rhetorical schema(See (Horn, 1991; Ward, 1990)). This will be discussed further in chapter 6.

2.5 Scalar Implicatures

I am not going to explain in detail the analysis of scalar implicature that I assume here(Hirschberg, 1985), hoping that a gloss will suffice to explain examples 26 and 27.

In the following excerpt, Doug (d) uses the term *girlfriend* in 26-9 which implicates, by the Quantity Maxim(Grice, 1967), that the woman he is speaking of is not his wife.

- (26) (7) d. I'm fine. I've got a general tax question for you.
 (8) h. sure
 (9) d. uh *my girlfriend* and I bought a house in december. And uh we're both on the mortgage. and we're not sure as to how to handle the mortgage, or excuse me how to handle the deductions.
 Uh every example I see in tax books usually cites a married couple, though and WE'RE NOT MARRIED. and we uh I'm afraid that this year she will not qualify for uh to itemize. I will – I'll be well over it.
 (10) h. Who made the payments?

However, a few utterances later, he makes this implicature explicit, by actually stating that they are not married. In this case the speaker makes his own implicature explicit. However in example 27, Maurie (m) in (15) requests Harry to make an implicature explicit that was implicated in Harry's utterance in (14).

- (27) (12) h. then they are remiss in not sending it to you because that money is taxable sir
 (13) m. i i know it's taxable, but but i thought they would wait until the end of the 30 months
 (14) h. no sir, *not for state purposes or federal purposes neither way, and remember for state purposes, you do not get a deduction for the interest you're paying em*
 (15) m. and for federal?
 (16) h. FOR FEDERAL PURPOSES YOU DO GET A DEDUCTION FOR THE INTEREST THAT YOU'RE PAYING
 (17) m. that's the uh that was my question okay

This implicature is what is actually conveyed in Harry's utterance at (16). The inference rule for implicature depends on the fact that (14) provides a salient scale, of *state purposes*, *federal purposes*. This scale is only partially ordered; in fact the two values are incomparable. Since (14) makes this scale salient, Harry's denial of *for state purposes, you get a deduction for the interest that you are paying* in (14), licenses a scalar implicature. This implicature is generated by the recognition of the denial of a proposition predicated on a member of a salient scale. The denial licenses the implicature that either Harry doesn't know whether the proposition can be denied for other members of the scale or Harry believes the proposition can't be denied, in other words the proposition holds. Therefore, since Harry is in a position to know, Harry has implicated that *for federal, you do get a deduction for the interest that you are paying*. Nevertheless the caller requests that this implicature be made explicit in (15).

2.6 Summary

This completes the discussion of the types of examples that are examined in this proposal. We have seen that IRUs can have various logical properties, and can be explicitly or implicitly related to their antecedents. They can vary in terms of their location, which can be adjacent or remote, or in the same speaker's last utterance. An IRU and its antecedent may also be within one turn of a speaker. Another conversant may have added the antecedent to the discourse situation, or the speaker of the IRU may be reformulating his own contribution. They may be explicitly embedded in a propositional operator such as a conditional. An IRU may be elliptical or a full clause, and may be said with falling or rising intonation. These properties will be discussed again in chapter 7. The next chapter describes a model of mutual beliefs that supports the analysis given here.

Chapter 3

Belief Model

In order to explain how the hypotheses will be developed, I must first characterize some properties of the model of mutual beliefs. This is because utterances operate on the mental state of the discourse participants. In other words, utterances are intended to affect cognitive representations. It seems then that an accurate characterization of utterance effects depends on an accurate characterization of how beliefs are augmented and revised.

The beliefs model for dialogue must distinguish the properties of various types of antecedents of IRUs that were discussed in section 2.2.2. These properties include whether the antecedent was entailed via logical inference, or whether it was implicated and thus defeasible. The model should also distinguish between beliefs that are made public and those that are not since what has been said is clearly part of the discourse situation. In contrast, what is inferable may not be part of the discourse situation due to agents' limited inferential capabilities. These properties have to do with the evidence supporting a belief.

It is also desirable to distinguish beliefs that are currently attended to from those that are not, in order to capture the differences between IRUs with salient antecedents and ones whose antecedents occurred much earlier in the dialogue. Whether or not a belief is currently attended to at a particular point in a dialogue is independent of the evidence supporting a belief. A fact may be believed very strongly, and yet not be currently in an agent's consciousness. Alternately an agent may be obsessed with a belief for which she has no evidence. I shall argue that being 'in consciousness' is a critical determinant of the ease with which inferences are made (Kahneman et al., 1982). Section 3.1 will discuss the evidence aspect of the beliefs model. Salience will be discussed in section 3.2.

3.1 Shared Environment Model of Mutual Beliefs

The beliefs model proposed here is based on Lewis's SHARED ENVIRONMENT model for common knowledge (Lewis, 1969; Clark and Marshall, 1981). In this model, mutual beliefs depend on evidence, openly available to the conversants, plus a number of underlying assumptions.

Shared Environment Mutual Belief Induction Schema

It is mutually believed in a population P that Ψ if and only if some situation \mathcal{S} holds such that:

1. Everyone in P has reason to believe that \mathcal{S} holds.
2. \mathcal{S} indicates to everyone in P that everyone in P has reason to believe that \mathcal{S} holds.
3. \mathcal{S} indicates to everyone in P that Ψ .

The situation \mathcal{S} , used above in the mutual belief induction schema, is the discourse situation. Condition (3) of the induction schema states that the situation must be public, and it must publicly indicate which agents of a population P have access to the situation. Condition (3) then specifies that mutual beliefs are indicated by a discourse situation. The discourse situation itself is characterized by three distinct notions: PUBLIC EVENTS, ASSUMPTIONS and ENDORSEMENTS.

The discourse situation consists of a sequence of public UTTERANCE EVENTS. Each public utterance event changes the discourse situation producing a new discourse situation. According to step (3), agents must determine what the discourse situation indicates in order to determine what is mutually believed, i.e. Ψ . In other words, agents must reason about what the discourse situation, \mathcal{S} , indicates, and this depends on interpreting a public event according to a set of ASSUMPTIONS. The fact that the inference of a particular mutual belief relies on a particular set of assumptions already must be mutually believed.¹ Which assumptions underlie a particular inference may be partially defined by the previous discourse situation, by conventions of communication, or by agents' shared background (Clark and Marshall, 1981; Perner and Garnham, 1988; Grice, 1957). The final piece is that each assumption has an associated ENDORSEMENT. This endorsement is meant, roughly, to indicate how strongly the assumption is believed.

An example of the inference rule schema for the model of mutual belief is given below. A Public-Event can be any action that is done openly in a shared environment; I will only consider utterance events in this proposal. The symbol \rightsquigarrow means inferable as a **mutual belief** under certain assumptions.² The assumptions are given with the rule and the [endorsement] annotation indicates the type of the endorsement that each assumption receives.

Public-Event(?A, ?B, ?U, ?P) \rightsquigarrow Fact(?B, ?P) [endorsement]

Assumptions = assumption-1 (?A, ?B) [endorsement]

..

..

assumption-n (?B ?U ?P) [endorsement]

¹Since the inference of mutual beliefs depends on a conversant's own reasoning about what a discourse situation indicates, each conversant must have their own model of what is mutually believed. Conversants work under the assumption that their models are in fact the same, and the process of conversation demonstrates their efforts to keep them the same.

²I do not intend to address the frame problem.

Examples using this schema will be given in the next section where I describe the model of how agents achieve mutual belief of understanding and acceptance. However, first I must discuss the nature of endorsements.

3.1.1 Endorsements

Endorsements provide a qualitative way of distinguishing between beliefs that an agent would easily give up and those that an agent would rarely change. An endorsement can be very weak, meaning that the assumption that has that endorsement is very weak, entailing that a mutual belief that depends on that assumption is easily defeasible. The types of endorsements are categorized and ordered from weakest to strongest: HYPOTHESIS < DEFAULT < ENTAILMENT < LINGUISTIC < ABSOLUTE (See (Prince, 1981b; Clark and Marshall, 1981; Galliers, 1991b)). This schema defines a model of mutual beliefs that can range from mutual suppositions to mutual knowledge (Prince, 1978; Nadathur and Joshi, 1983).

The ordering on the types of endorsements reflects the **relative** defeasibility of different assumptions; an assumption endorsed as a DEFAULT can get defeated by LINGUISTIC information. For example, suppose that the mutual belief is *Madison can swim*. This belief is based on two shared assumptions: (1) *Madison is a dog*, (2) *Dogs can swim*. Let's say assumption (1) is something that two agents can see with their own eyes, so it might be strongly endorsed, e.g. ABSOLUTE. In the absence of other evidence, agents make a default inference about dogs and their ability to swim. Assumption (2) is thus endorsed as a DEFAULT. Since (2) is only a default, it can be easily defeated. If Madison's owner comes along and tells the two agents that *Madison can't swim*, they are likely to abandon their mutual belief that Madison can swim, i.e. the mutual belief that *Madison can swim* is defeated. It is also possible to upgrade the endorsement of an assumption, and thereby make it less defeasible. For instance, Madison's owner might come along and tell the two agents *Madison is a very good swimmer*. The mutual belief that *Madison can swim* now has an endorsement type of LINGUISTIC. One of the roles of IRUs that will be discussed in the next section is this function of upgrading the endorsement for a mutual belief.

Endorsements can provide the distinctions between different types of information that were discussed in sections 2.3, 2.4 and 2.5. The claim is that some types of endorsements provide better support for beliefs than other types. The type HYPOTHESIS is the weakest possible endorsement. It is used to endorse assumptions that have no evidence supporting them at all, and DEFAULT is used for defeasible inferences such as implicatures and what I will call conversational defaults in section 4.1. The endorsement ENTAILMENT is used for entailments and presuppositions that have not been made explicit. LINGUISTIC endorses assumptions that have been made explicit in the dialogue and ABSOLUTE refers to assumptions that are incontrovertible as though from divine authority. Of course the same assumption can have multiple endorsements; it may be both entailed by what has been said as well said explicitly, i.e. it has endorsements of both types ENTAILMENT and LINGUISTIC.

A claim of this proposal is that Attitude and Consequence IRUs function to upgrade the type of an endorsement on an assumption underlying a mutual belief. This then makes the mutual belief that depends on that assumption less defeasible.

As a simple example of this claim, consider the fact that whether an utterance event has any effect at all depends on an assumption that the addressee actually heard what was said. Let's call this the COMPLETE HEARING assumption (Clark and Schaefer, 1989). In many cases, explicit evidence that would support this assumption is never provided in the discourse situation. However an addressee may repeat verbatim what was just said as in 28 below:

- (28) A: The number is 427 899.
 B: 427 899.

B's repetition means that the complete hearing assumption now has an endorsement type of linguistic. This claim will be discussed further in section 4.1. First though, I must briefly digress to address the question of how the endorsements on assumptions as given in the schema above, combined with the endorsement on the inference rule used to infer $\text{Fact}(\text{?B}, \text{?P})$, affect the strength with which $\text{Fact}(\text{?B}, \text{?P})$ is believed.

3.1.2 Combining Endorsements

In order to address the problem of how it is that the endorsements on assumptions are combined and reasoned about, for the time being, I adopt a rule that a chain of reasoning is only as strong as its weakest link.

Weakest Link Rule: The endorsement of a belief P depending on a set of underlying assumptions a_i, \dots, a_n is $\text{MIN}(\text{Strength}(a_i, \dots, a_n))$

This seems intuitively plausible and means that the endorsement of a belief depends on the endorsements of the underlying assumptions. It also means that for all inference rules that depend on multiple assumptions, the endorsement of an inferred belief is the weakest of the supporting beliefs. Since the inference rule itself is one of the supporting beliefs, and it has an associated endorsement, this means that some kinds of inferences, such as implicatures, cannot be believed as more than defaults no matter how strong the endorsements are on the assumptions.

This account of mutual beliefs has been strongly influenced by Gallier's account of single agent beliefs (Galliers, 1991b). According to Galliers, the strength of a belief depends on both the number of supporting assumptions as well as their individual endorsements. The number of supporting assumptions gives a measure of the coherence of the whole belief set (Harman, 1986). Coherence of a set of beliefs is determined by links of 'explanation' and 'causality'. This view is cognitively plausible, yet it raises problems with the determination of the relative weight of these two factors. It also means that the granularity of the assumptions underlying a belief can affect the calculation of how strongly a belief is endorsed. This aspect of Gallier's system is still under development, and I am avoiding these issues through the use of the weakest link rule.

3.2 Salience of Beliefs

There are two points I wish to make about modeling the salience of a belief. The first point is that salience is a binary distinction that cuts across all the types of beliefs discussed earlier. The salience of a belief is completely independent of the assumptions underlying it and the endorsements on these assumptions. Salience is not an endorsement type for beliefs, rather any type of belief can be currently salient or not.

The second point is that there seems to be evidence that the proposition realized by the most recent utterance has a special status in terms of salience (Anderson and Bower, 1973; Landauer, 1975; Clark and Clark, 1977). I will assume the most recent utterance corresponds to the most salient proposition. For convenience I will refer to most recent utterance as the *CENTERED* proposition, following work on referential discourse entities (Grosz et al., 1983; Grosz et al., 1986). This recency-based definition of salience doesn't account for sentential and discourse subordination. Nor does it account for the effect of frequency on discourse salience. Thus this definition is a temporary working definition that will be used in the description of examples from the corpus discussed in section 7. The attention/memory model will be refined as part of the proposed work. This will permit the exploration of specific hypotheses about the effects of frequency on salience in the course of performing the dialogue simulation experiments proposed in chapter 8.

In computational terms, therefore, it is relatively simple to model whether beliefs are salient; either they are part of a special subset of beliefs in the current focus space, or they are not. The more difficult question is how beliefs get moved into and out of the current focus space. This will be discussed in chapters 6 and 8.

3.3 Types of Propositions

Propositions can be of many different types: events, acts, processes, states (Webber, 1978). At the risk of oversimplifying, I will make a simple distinction between propositions that describe states and facts, and those that describe actions or goals. I will call factual propositions belief-props, and those that describe actions act-props.

This distinction is necessary because agents in advisory dialogues discuss beliefs about the current state as well as future actions and the effects of these actions.³ When I discuss accepting an utterance in chapter 4, there is a distinction dependent on whether the utterance realizes a belief-prop or an act-prop. If the propositional content of an utterance is a fact, acceptance of the utterance means believing the propositional content. If the propositional content proposes a future action of the agent, then acceptance may generate an intention to act. At the least, acceptance will generate the consideration of adopting an intention to act. This distinction will be discussed in sections 5.1 and 8.2.

³I am focusing on dialogues where the execution of actions is not interleaved with the discussion. All actions are to be executed in the future.

3.4 Summary

This representation of mutual belief differs from the common representation in terms of an iterated conjunction (Cohen and Levesque, 1990; Litman and Allen, 1990) in that: (1) it distinguishes between private beliefs and public beliefs; (2) it allows one to represent the different kinds of evidence for mutual belief; (3) it controls reasoning when discrepancies in mutual beliefs are discovered since the assumptions and their endorsements can be inspected; (4) it consists of a finite decision procedure of checking the three conditions in the mutual belief induction schema rather than an infinite list of statements. The next chapter will present a model of how mutual understanding and acceptance are achieved that depends on this model of mutual belief.

Chapter 4

Attitude

I claimed in section 1.2 that agents in a dialogue need to achieve communicative functions related to Attitude. They need to achieve mutual belief in the understanding and acceptance of an utterance in order to have a collaborative plan involving the proposition realized by that utterance. Yet while understanding and acceptance can be communicated directly through utterances such as *I understand* and *I agree*, such explicitness is very rare. Agents must therefore infer whether in fact mutual understanding and acceptance have been achieved.

The key claim of this chapter is that agents monitor the effects of their utterance actions in order to make these inferences. However, even if an agent monitors for the effect of her utterance, she cannot directly perceive the effect of an utterance because she does not have direct access to another agent's mind. Therefore, she must infer the effect of the utterance by observing the other agent's behavior.

There are two key points about how these inferences are made: (1) there must be some public behavior by the addressee that licenses the inference, and (2) this public event must occur soon after the original event in order for dialogue to operate efficiently.

I claim that the way that mutual understanding and acceptance are achieved is through a two part process: (1) A speaker produces an utterance. This public event weakly licenses a mutual belief of understanding and acceptance as a HYPOTHESIS, irrespective of the good intentions of the speaker. (2) The next action by the addressee is taken as evidence of the effect of the speaker's utterance.¹

This evidence is then reasoned about. Both conversants are cognizant of the fact that this is how conversation works, so an addressee, when planning her utterance, in addition to planning its content, must also reason about what it will demonstrate about the effect of her co-conversant's last utterance.

This is a local process, i.e. it is the next utterance that is monitored for the effect of the previous. This has been characterized as a convention or as a 'rule of conversation' (Sacks et al., 1974). But rather than being merely conventional, it seems that there are very good functional reasons for the convention to exist.

¹Except for circumstances where it is clear that the flow of the conversation has been interrupted.

First, consider psychological studies about the limits on attention/memory. These studies found that the verbatim content of an utterance is retained for a very short period of time (Bransford et al., 1972; Anderson, 1974). This means that any action in dialogue that has a communicative function related to the understanding and acceptance of a proposition communicated by an utterance, is constrained to occur very soon after that utterance, or risk incurring an additional cost of reinstantiating that utterance as the current locus of attention.

Furthermore, there is an abundance of other psychological data on debriefing (Kahneman et al., 1982), showing that inferences based on facts that are untrue will persist even when the original fact has been retracted. This is incorporated in Harman's Principle of Positive Undermining (Galliers, 1990; Harman, 1986):

Principle of Positive Undermining:

Only stop believing a current belief if there are positive reasons to do so, and this does not include an absence of justification for that belief.

This means that in order to retract all beliefs that were added as an inference from a false belief, each inferred belief would independently have to be explicitly challenged. In combination with the fact that belief revision is a costly operation, these observations show that there is a strong pressure on conversants to ensure that they have understood and been understood in the intended way **at the time** the propositional content of the utterance is initially added to the representation of mutual beliefs. In other words, there is strong motivation for **local management** of potential misunderstandings and disagreements and for **explicit signals** by conversants as to their beliefs about the current propositional content.

This means that all of the IRUs related to Attitude occur soon after their antecedents. I will focus on a special case of this, those cases where the IRU is adjacent to its antecedent. The way in which communication goals of Attitude are achieved is discussed in the remainder of this chapter.

4.1 Mutual Belief of Understanding

This section considers the inference of mutual understanding. Note that what is being modelled here is mutual belief that the utterance has been **understood as intended**, not mutual belief in the truth of the propositional content of the utterance. That will be discussed in section 4.2.

I adopt the assumption that the participants in a dialogue are trying to achieve some purpose (Grosz and Sidner, 1986) and that some aspects of the structure of dialogue arises from the structure of these purposes and their relation to one another. The minimal purpose of any purposeful dialogue is that an utterance be understood, and this goal is a prerequisite to achieving other goals in dialogue, such as another agent believing the proposition conveyed, or committing to a future action described by that proposition. Thus achieving **mutual belief of understanding** is an instance of the type of activity that agents must perform as they collaborate to achieve the purposes of the dialogue. I expect that the details of the model that

are needed to support the achievement of mutual belief of understanding can be extended to the achievement of other goals in dialogue.

Utterance intentions are not made explicit and the effects of an utterance on the addressee are not transparent to the speaker. This means that achieving understanding is not unproblematic; it is a process that must be managed, just as other goal achieving processes are (Clark and Schaefer, 1989; Brennan, 1990). While the speaker may have their own private beliefs about whether or not the utterance will have the intended effect, this description is meant to characterize what is made public by an utterance event. All that is made public is that the speaker has a hypothesis that this utterance will achieve some effect. I will argue that it is the addressee's actions that offer evidence to the speaker about the effects of the utterance.

According to the shared environment model, an utterance event in a discourse situation licenses the inference of mutual understanding, under certain assumptions.² Mutual belief that the addressee B understood U to mean P, is based on assumptions such as: (1) the addressee B is attending to the utterance (the ATTENTION assumption); (2) the addressee hears U (the COMPLETE HEARING assumption); and (3) the addressee believes that U realizes the proposition P that the speaker intended to convey (the REALIZATION assumption).³ This is depicted in the inference rule schema given below. In this schema, the term $\text{say}(A, B, U, P)$ describes a public utterance event in which an agent A says an utterance U to another agent B meaning to convey a proposition P.⁴ The assumptions described above are given with the rule.

$\text{say} (?A, ?B, ?U, ?P) \rightsquigarrow$	$\text{understand} (?B, ?P)$	[absolute]
Assumptions =	$\text{attend} (?B, ?U)$	[hypothesis]
	$\text{hear} (?B, ?U)$	[hypothesis]
	$\text{believe} (?B, \text{realizes} (?U, ?P))$	[hypothesis]

An utterance demonstrates a public hypothesis that it will achieve its purpose. This public event licenses the inference of mutual belief according to the mutual belief induction schema. Each of the assumptions underlying the mutual belief starts out with an endorsement of HYPOTHESIS. It isn't until **after** the addressee's next action that an assumption may have its endorsement modified. Since assumptions are very defeasible when they are only hypotheses, this representation allows further events in the situation to cancel the inference of mutual understanding. For example the addressee may say *What?* which defeats the COMPLETE HEARING assumption. However, further events in the discourse may also **upgrade** the endorsements on the assumptions underlying the inference of mutual understanding. For instance, the way some IRUs function to achieve mutual belief of understanding is by addressing the assumptions underlying the inference of mutual understanding. For instance, a repetition directly addresses the complete HEARING assumption.

Sections 4.1.1 and 4.1.2 give examples of repetitions and paraphrases from the HG corpus.

²Of course a single utterance event may license the inference of a mutual belief in other facts as well as inference of mutual understanding.

³Of course the addressee may believe that U realizes some other proposition R.

⁴Again I stress that I do not mean that P is conveyed directly and I use the realization assumption to represent this fact (Reddy, 1979).

Examples of making inferences and implicatures explicit will be discussed in sections 4.1.3 and 5.1. Section 4.2 will discuss the inference of mutual acceptance.

4.1.1 Example of a Repetition

Consider example 2 from section 1.1, repeated here for convenience:

- (29) (6) r. uh 2 tax questions. One: since April 81 we have had an 85 year old mother living with us. Her only income has been social security plus uh approximately three thousand dollars from a certificate of deposit. and I wonder what's the situation in terms of claiming her as a dependent, or does that income from the certificate of deposit rule her out as a dependent?
 (7) h. Yes *it does*
 (8) r. IT DOES
 (9) h. Yup that knocks her out.

Ray, in (8), repeats Harry's assertion from (7). This upgrades the endorsement on the hearing and attention assumptions associated with utterance (7) from HYPOTHESIS to LINGUISTIC. The realization assumption, i.e. what proposition p7 is realized by u7, remains endorsed as a DEFAULT because Ray merely repeated what he heard and a repetition doesn't provide any evidence that Ray actually understood what Harry meant. This instantiates the inference rule for understanding as follows:

say(harry, ray, u7, p7) \rightsquigarrow	understand(Ray, u7, p7)	[absolute]
Assumptions =	attend(Ray, u7)	[linguistic]
	hear(Ray, u7)	[linguistic]
	bel(Ray, realize(u7, p7))	[default]

Because of the WEAKEST LINK rule, the mutual belief about understanding is still a default. That is, the conclusion of the inference process is

MB(understand(Ray, u7, p7)) [default]

However, the attention and complete hearing assumptions are no longer defeasible by linguistic evidence. This means that it would be infelicitous for Ray later on to say *Oh I didn't hear you say that* or *I thought you said that it doesn't*. The next section examines the effect of a paraphrase on the inference of mutual understanding.

4.1.2 Example of a Paraphrase

Consider example 1 from section 1.1, repeated here as 30:

- (30) (18) h. I see. *Are there any other children beside your wife?*
 (19) d. *No*
 (20) h. YOUR WIFE IS AN ONLY CHILD
 (21) d. right. and uh wants to give her some security

Harry's utterance of (20) is said with a falling intonational contour and hence is unlikely to be a question. This utterance results in an instantiation of the inference rule as follows:

say(harry, ray, u20, p20) \leadsto understand(Ray, u20, p20) [absolute]
 Assumptions = attend(Ray, u7) [linguistic]
 hear(Ray, u7) [linguistic]
 bel(Ray, realize(u7, p7)) [linguistic]

Because of the WEAKEST LINK rule, the mutual belief about understanding is now endorsed as linguistic.

MB(understand(Ray, u7, p7)) [linguistic]

The belief in the achievement of mutual understanding is licensed by a LINGUISTIC endorsement since all of the underlying assumptions are endorsed as linguistic. Thus a paraphrase provides excellent evidence that an agent actually understood what another agent meant.

4.1.3 Making Entailments Explicit

While Consequence is given as a separate function from Attitude, there is actually an overlap in their functionality. I will describe here the set of cases where inferences are made explicit when they are adjacent to at least one of their antecedents. It should be noted however that these cases function communicatively for both Consequence and Attitude.

Cases where a speaker A, intends an addressee B, to infer Q, from an utterance that realizes a proposition P, can be represented by an additional assumption on the understanding inference rule schema. The fact that A intends B to infer Q, by saying U, which realizes P, in a particular discourse situation, can be represented by an additional condition that P licenses Q in the situation as given below:

say(?A, ?B, ?U, ?P) \leadsto understand(?B, ?Q) [absolute]
 Assumptions = attend(?B, ?U) [hypothesis]
 hear(?B, ?U) [hypothesis]
 believe(?B, realize(?U, ?P)) [hypothesis]
 believe(?B, licenses(?P, ?Q)) [entailment]

This new fifth assumption will be called the LICENSE assumption. As an example, consider example 31, repeated from section 2.3, where Harry makes an inference explicit in 31-(17). This inference follows from modus ponens and the previous two utterances.

- (31) (15) h. oh no. IRA's were available as long as you are not a participant in an existing pension
 (16) j. Oh I see. well I did work I do work for a company that has a pension
 (17) h. ahh. Then you're not eligible for eighty one

According to the model of understanding outlined in section 4.1, extended for inferences above, the fact that this proposition was inferable after (16), would be represented by the license assumption.

believe(H, license(p16, p17)) [entailment]

The license assumption providing support for whether the inference is mutually believed is endorsed as an ENTAILMENT. However utterance (17) upgrades the endorsement on this assumption to LINGUISTIC.

believe(H, license(p16, p17)) [linguistic]

Making an inference explicit also upgrades the ATTENTION, COMPLETE HEARING, and REALIZATION assumptions. Any utterance upgrades attention. The COMPLETE HEARING, and REALIZATION assumptions get upgraded because it isn't possible to make a valid inference from an utterance unless one heard it and understood its propositional content correctly.

4.1.4 Summary

Each type of IRU, the assumption addressed and the endorsement provided is given in Figure 4.1. I consider the inferences about what is mutually believed that are licensed by the addressee's utterance as implicatures that arise from norms of interaction. They have the properties of implicatures of being both reinforceable and defeasible (Sadock, 1978). I will refer to these inferences as CONVERSATIONAL DEFAULTS (Joshi et al., 1986). Examples of the kind of behavior that licenses these inferences were provided in sections 4.1.1, 4.1.2, and 4.1.3.

As detailed in Figure 4.1, prompts, repetitions, paraphrases and making entailments explicit all provide a linguistic endorsement on the ATTENTION assumption. The only thing that prompts, such as *uh huh*, do directly is to provide a linguistic endorsement for attention.

However repetitions, paraphrases and making entailments explicit also upgrade the endorsement on the COMPLETE HEARING assumption. This is represented by upgrading the endorsement for the complete hearing assumption to linguistic. A repetition demonstrates complete hearing of the verbatim content of what was said (Clark and Brennan, 1990). A paraphrase demonstrates complete hearing by showing that the verbatim content has been semantically incorporated into the addressee's memory. Making an entailment explicit demonstrates complete hearing by showing that the verbatim content has been incorporated into the addressee's memory and that at least one inference has been performed on this content.

NEXT Utterance Type	ASSUMPTION ADDRESSED	ENDORSEMENT TYPE
PROMPT	attention	linguistic
REPETITION	attention, hearing	linguistic
PARAPHRASE	attention, hearing, realize	linguistic
ENTAILMENT	attention, hearing, realize, license	linguistic
IMPLICATURE	attention, hearing, realize, license	linguistic
ANY Next Utterance	attention, hearing, realize, license	default

Figure 4.1: How the Addressee's Following utterance upgrades the endorsement for assumptions underlying the inference of mutual understanding

In addition, a paraphrase and making an entailment explicit provide a linguistic endorsement for the *REALIZATION* assumption, of what proposition the paraphraser believes the previous utterance realizes. Paraphrases do this by demonstrating that the content has been semantically incorporated into memory. Explicit entailments additionally provide evidence of what inferences the inferrer believes the realized proposition licenses in this situation.

In addition, **any** next utterance by the addressee can upgrade the endorsements of the underlying assumptions to default. (See Figure 4.1). Of course a default endorsement is weaker than a linguistic endorsement. The basis for these default inferences will be discussed in section 4.2.

In each case, the IRU addresses one or more assumptions that have to be made in order to infer that mutual understanding has actually been achieved. The assumption, rather than endorsed as a hypothesis or a default, gets upgraded to an endorsement type of linguistic as a result of the IRU. The fact that different IRUs address different assumptions leads to the perception that some IRUs are better evidence for understanding than others, e.g. a paraphrase is stronger evidence of understanding than a repetition (Clark and Schaefer, 1989).

The model of beliefs including assumptions and endorsements is implemented in a modified Truth Maintenance System (TMS) based on Galliers' theory of Autonomous Belief Revision (ABR) (de Kleer, 1986; Cawsey et al., 1992). This will be discussed further in chapter 8 where the computational simulation is described.

Other analyses of the function of these adjacent IRUs are possible. Indeed, they may function to maintain a proposition as salient as much as provide evidence of understanding. This is

actually supported by the fact that of all the adjacent IRUs in the corpus, **none** of them are matched by examples where the addressee actually got it wrong. In other words, there are no cases of attempts at adjacent repetitions or paraphrases that demonstrate that the addressee misunderstood what the speaker had said (Tannen, 1989). This highlights a potential overlap in the Attitude and Attention functions which will be explored in the simulation experiments described in chapter 8.

4.2 Acceptance and Rejection

In section 4.1, I suggested that the assumption that an utterance always achieves the intended effect, should be replaced by adopting a model in which agents' behavior provides evidence for whether or not mutual belief has been achieved. I used the achievement of mutual belief of understanding as an example of how monitoring the evidence supplied by the next utterance can license the inference of mutual belief. I argued that local management of these processes is more than a convention; local management lets resource-bounded agents achieve mutual belief efficiently.

But in many discourse situations agents want to do more than achieve mutual understanding. It is important to distinguish an agent actually **ACCEPTING** the belief that P or intending to perform an action described by P, from merely **UNDERSTANDING** that P was conveyed. Some accounts legislate that helpful agents should adopt others' beliefs and intentions (Cohen and Levesque, 1990; Litman and Allen, 1990). A more sophisticated approach using default logic maintains that acceptance depends on whether or not the agent previously believed $\neg P$ (Perrault, 1990; Grosz and Sidner, 1990). These are simplifying assumptions: agents are autonomous and can decide whether, as well as how, to revise their beliefs (Galliers, 1991b).

Achieving understanding and compensating for resource-bounds are issues for a model of dialogue whether or not agents are autonomous. However agents' autonomy means there are a number of other reasons why A's utterance to B, intending to convey P, might not achieve its intended effect: (1) P may not cohere with B's beliefs; (2) B may not think that P is relevant; (3) B may believe that P does not contribute to the mutual goal; (4) B may prefer doing or believing some Q where P is mutually exclusive with Q, (5) if P is about an action, B may want to partially modify P with additional constraints about how, or when P.

Yet, as I noted earlier, it is rarely the case that explicit evidence of agreement is given by an utterance such as *I agree*. And, typically, the same forms that are used to convey simply understanding, may convey agreement (but see (Hockey, 1991)). So how do agents achieve mutual belief of acceptance when acceptance is necessary? My hypothesis is that a mutual belief that 'acceptance is necessary' determines how this process works. Just as goals constrain the way instructions are understood and evaluated (Pollack, 1986; Di Eugenio, 1992), mutual goals constrain whether an utterance licenses the inference of mutual acceptance. It must be mutually believed that acceptance of the utterance is necessary to achieve the mutual goal in order for an inference of acceptance to be licensed. When this situation holds, acceptance can be inferred via the operation of a simple principle of cooperative dialogue.⁵:

⁵This is a simplification of the COLLABORATIVE PLANNING PRINCIPLES described in (Whittaker and Stenton,

COLLABORATIVE PRINCIPLE: Conversants must provide evidence of a detected discrepancy in belief as soon as possible.

The collaborative principle is an instance of a general rule incumbent on a cooperative conversant to prevent his/her co-conversant from making false inferences (Joshi, 1982; Joshi et al., 1986). This principle claims that evidence of conflict should be made apparent in order to keep DEFAULT inferences about acceptance or understanding from going through. Without it, a conversant might not bring up an objection until much later in the conversation, at which point the relevant belief and some inferences following from that belief will have been added to the common ground as DEFAULTS. The retraction of that belief may result in many beliefs having to be revised. Or worse, according to the Principle of Positive Undermining (Harman, 1986; Galliers, 1991b), some of these faulty beliefs may never be successfully retracted.

The principle claim of this section is that there is a large subset of IRUs that cannot function as evidence for conflicts in beliefs or intentions via their propositional content, i.e. they cannot function as evidence of rejection of another speaker's utterance. This follows fairly directly from the fact that they are informationally redundant, but there are some additional constraints: (1) they are adjacent to their antecedent; (2) the other conversant said the antecedent; (3) they are said with broad focus; (4) they are said with a falling boundary tone; (5) they must include the information focus of the antecedent if the IRU is a repetition. This will be discussed further below.

As an example of the inference of mutual acceptance, consider Mary's utterances of *yes*, *alright* or *I see* in 32-(47), (49), (53), and (55). According to the proposal here, these cannot function to reject Harry's previous assertion, nor can Harry's *ok* in (50) and (56).

- (32) (46) h. Ok, the remaining roughly ninety thousand
 (47) m. Yes
 (48) h. I would like to see that in two different, municipal bond funds, that are variable rate funds
 (49) m. Yes
 (50) h. Ok
 (51) m. Heard you telling another person that a couple of calls back
 (52) h. Right. There are a number of variable rate municipal bond funds floating around. Contact your broker and ask him to recommend a couple of them
 (53) m. Alright
 (54) h. And they will maintain their value approximately because they are variable rate funds
 (55) m. I see
 (56) h. Ok
 (57) m. Fine

Since in this situation, I assume that Harry and Mary have a mutual goal of Mary accepting the advice that Harry offers to address her problem, the inference of mutual acceptance is licensed

1988; Walker and Whittaker, 1990).

by the fact that Mary doesn't provide evidence of rejection when she has an opportunity to do so.

Some researchers have argued that PROMPTS have propositional content of their own, such as *I agree* (Grosz and Sidner, 1986; Cohen et al., 1990; Levesque et al., 1990). On these accounts it is necessary to provide such explicit evidence of agreement or understanding. But the account here is that acceptance is inferred from the fact that nothing was said to indicate non-acceptance. One way in which non-acceptance can be recognized is discussed in the next section.

4.2.1 Implicit Rejection and Given/New

I've claimed that adjacent IRUs always indicate acceptance, but I've limited this claim by a number of constraints on these IRUs. These constraints were given in section 4.2. One of the constraints, (5), was that the IRU must include the information focus of its antecedent. This section examines cases where constraint (5) is violated since it illustrates a general fact about the nature of demonstrating acceptance. It also shows that inferring a rejection depends on the given-new structure of the utterances in a dialogue.

Hirschberg demonstrated how disagreements (rejections) can be inferred as a special case of scalar implicatures (Hirschberg, 1985). This inference depends on identifying two utterances as differing by alternate values of a scale. Generally this means that these utterances will differ in what the new information is in the second utterance, and thus one factor in recognizing the scalar implicature is identifying the new information in an utterance (Rubinoff, 1987). However, the phenomenon seems to be slightly more general than this, as captured by condition (5). Consider the following example from (Levinson, 1979):

- (33) A: There's a man in the garage.
B: There's something in the garage.

Utterance 33B violates condition (5). In this case the propositional content of B's utterance is an entailment of A's utterance, B's utterance meets conditions (1) through (4) above for IRUs providing evidence of understanding and acceptance, and yet B's utterance actually serves as a rejection of A's. I claim that the reason it does so is because it does not realize the information focus of A's utterance. Consider the attested example in 34:

- (34) A. We bought these pajamas in New Orleans for me.
B. We bought these pajamas in New Orleans. (LW 5/24/92)

A's utterance was said with the phrase *for me* marked as new information, the information focus. When B repeats all of the utterance except for the information focus, B implicates that B rejects the main predication of A's utterance.

This means that a representation of utterance meaning for dialogues between autonomous agents will have to represent and reason about given and new information in order to accurately

represent the way non-acceptance functions in naturally occurring dialogues. The information that will license the inference of rejection cannot be coded as logical inconsistency alone. There is no sense in which the B utterances in 33 and 34 are logically inconsistent with the A utterances. However, the inference of rejection is beyond the scope of this thesis. I merely note here that it would require access to information of three types: (1) logical inconsistency; (2) the notion of alternate members of a set (Hirschberg, 1985); and (3) Given/New structure.

Chapter 5

Consequence

The second set of hypotheses about the function of IRUs involve Consequence: to augment the evidence supporting mutual beliefs that certain inferences are licensed, or to provide a context in which an inference is more easily made. What do I mean by the distinction between these two cases?

Consider P , an antecedent of an inference rule, Q a consequent, and an inference rule that can be expressed roughly as $P \rightarrow Q$. There are two states that an agent might be in with respect to making an inference. An agent may:

1. Believe P and Believe $P \rightarrow Q$, and have to infer Q .
2. Believe P and Believe Q , and have to infer $P \rightarrow Q$.

In both cases the inference can be made explicit. Examples of the first case, where Q is stated, were discussed as making entailments and implicatures explicit in sections 2.3 and 2.5. I characterized these as upgrading the endorsement on the `LICENSE` assumption and gave an example in section 4.1.3. The inferences that I considered were limited to those motivated by inference rule schemas like *modus ponens*, or linguistically based inferences like scalar implicatures, where the information on which the inference depends is provided in the dialogue.

I have not yet discussed examples of the second case where an agent must infer $P \rightarrow Q$. These will be discussed below. The point of examining different types of inferences is that a goal of the proposed work is to determine the following: under what circumstances is it beneficial for an agent to make an inference explicit or perform some utterance action that functions to make it ‘easier’ for another agent to make an inference? Determining the answer to this question requires a specific proposal about agents’ resource-bounded reasoning mechanisms. I will leave the description of this mechanism until chapter 8. Here I discuss the second type of inference further.

The second case is complicated by the fact that the symbol \rightarrow actually refers to a number of different relations that can hold between propositions that describe actions or beliefs. These

relations are often characterized as simply being modus ponens relations, but agents in all probability make much finer distinctions. Taxonomies have been proposed for these distinctions as ‘propositional relations’ or ‘coherence relations’(Hobbs, 1979; Cohen, 1987). Determining whether a relation holds sometimes depends on information about the domain, but it is the linguistic context that licenses the inference of these relations. At times a speaker may explicitly provide clues about what inferences are intended through the use of cue words such as *because*. However, it seems that some cases of IRUs function to make it ‘easier’ for an agent to infer these propositional relations.

I adopt a small set of propositional relations that are adequate for my purposes: **CONTRIBUTE**, **SUPPORT** and **WARRANT**. This is a very rudimentary set of relations and each relation could be made more precise(Hobbs, 1979; McKeown, 1985; Di Eugenio, 1992; Balkanski, 1990). However, the purpose of modeling the process of achieving a collaborative plan doesn’t require precise versions of these relations.

I use the **CONTRIBUTE** relation to describe a relation between two actions P and Q, where the performance of P leads to the achievement of Q (Grosz and Sidner, 1986). This very roughly means that performing P may enable performing Q, as *laying the carpet* **CONTRIBUTES** to *furnishing the house*. The **CONTRIBUTE** relation also relates actions where performing P results in having performed Q, e.g. *flipping the switch* results in *turning on the light*(Pollack, 1986).

The relation **SUPPORT** describes a relation between two beliefs where a belief in one supports a belief in the other(Grosz and Sidner, 1986). This characterizes the cases I have discussed of inferences where a belief Q is inferred from a belief P. In example 24, let P be *Joe had a pension plan for 1981* and Q be *Joe is not eligible for an IRA for 1981*. Then P **SUPPORTS** Q.

Finally, I introduce a new relation called **WARRANT** to describe a relation between a belief and an intention, where the belief is meant to be a reason for adopting an intention. For example your belief that *chocolate is good for you* may warrant adopting an intention to *eat chocolate*. The **WARRANT** relation is needed since agents are autonomous and reason about which goals they wish to adopt, rather than being given goals(Pollack, 1991).

In terms of the two types of propositions discussed in section 5.1, these three relations describe three different combinations of propositions: (**CONTRIBUTE** act-prop act-prop); (**SUPPORT** bel-prop bel-prop); and (**WARRANT** bel-prop act-prop). The previous sections gave examples of the **SUPPORTS** relation between two beliefs. The remainder of this section will give examples of the **WARRANT** and **CONTRIBUTE** relations.

5.1 Inferring Propositional Relations

This section describes a number of examples of IRUs to motivate the second aspect of the Consequence function of IRUs, the inference of propositional relations. Certain instances of IRUs seem to be said just so that a propositional relation can be inferred. The inference of these relations seems to depend on an IRU realizing a proposition P, occurring sequentially very close to an utterance realizing a proposition Q, where a propositional relation holds between P and Q(Hobbs, 1985; Hobbs and Martin, 1987). For example consider the following excerpt

that was part of a discussion about where to eat lunch:

- (35) (1) Listen to Ramesh.
(2) HE'S INDIAN. (DH, Nov 5, 91)

The discussion involved a mutual goal to instantiate *X* in the proposition *We should eat at restaurant X of type Indian*. Doing the action given in (1) is meant to CONTRIBUTE to this mutual goal, and the addressees were meant to infer a propositional relation between (1) and (2). The proposition in (2) *Ramesh is Indian*, which all the conversants already believed, was intended to serve as a WARRANT for adopting the goal given in (1) (Webber and Joshi, 1982).

Similarly, consider the following dialogue, repeated from section 2.3.1. Here Mary (m) tells Harry (h) that she and her husband are both retired in (43), and a number of other facts about their financial situation in successive dialogue from (43) to (45).

- (36) (42) h. Now what is your income situation
(43) m. *We're both retired* and our income for the year is about um 24 about 26 thousand
(44) h. Have you other securities than stock? have you any bonds or certificates?
(45) m. Yes yes we do. We have some certificates oh about uh 15 - 20 thousand, not much we're not rich - and we have a house completely paid for, have some land in the poconos, completely paid for - and uh that's actually the extent of our uh
(46) h. Ok. On the proceeds of that gm stock
(47) m. yes
(48) h. I'd like to see you put that into two different southern utilities
.
(clarification of southern utilities)
.
(58) h. and those I think that you will find that that will give you a good return, YOU ARE RETIRED and that is primarily what you are looking for, addition I think that eventually those stocks will rise as well.....

Beginning with (46) Harry suggests a course of action. In (58) he tells Mary something that she told him in (43), *you are retired*. However Harry's statement in this context leads to an inference of a warrant relation between that fact and the proposed course of action. In other words, Harry is implying that *X is retired* warrants *X wants a good return on her investment*. More precisely let:

A = Invest proceeds of GM stock in 2 different southern utilities
P = You are retired
G = Get a good return on your investment

Then what Mary is supposed to infer is that: (1) P WARRANTS G, and (2) A CONTRIBUTES to achieving G. Since Mary believes P, she may adopt G as a goal, and then intend to do A to contribute to achieving G.

Finally consider another case of making an entailment explicit, non-adjacent to the entailment's antecedent(s). In the following segment, Jane (j) describes her financial situation to Harry (h) and a choice between a settlement and an annuity.

- (37) (1) j. Hello Harry, my name is Jane
 (2) h. Welcome Jane
 (3) j. I just retired december first, and in addition to my pension and social security, I have a supplemental annuity
 (4) h. Yes
 (5) j. which I contributed to while I was employed
 (6) h. right
 (7) j. from the state of New Jersey mutual fund. and I'm entitled to a lump sum settlement which would be between 16,800 and 17,800 or a lesser life annuity. and *the choices of the annuity um would be \$125.45 per month.* that would be the maximum with no beneficiaries
 (8) h. You can stop right there: take your money
 (9) j. Take the money
 (10) h. Absolutely. YOU'RE ONLY GETTING 1500 A YEAR. at 17,000, no trouble at all to get 10 percent on 17,000 bucks.

Harry interrupts her at (8) since he believes he has enough information to suggest a course of action, and tells her *take your money*. To provide a WARRANT for this course of action he produces an inference that follows from what she has told him in (7), namely *You're only getting 1500 (dollars) a year*.

Presumably Jane would have no trouble calculating that \$125.45 a month for 12 months amounts to a little over \$1500 a year. However the juxtaposition of this fact against the advice to *take the money* licenses the inference that the fact that *she is only getting 1500 dollars a year*, is a WARRANT for adopting a goal to *take the money*(Levinson, 1979).

This adoption of the advice relies on the assumption that agents reason about the utility of performing certain actions as opposed to others. Agents deliberate with a general rule that *if one course of action, A, produces greater benefits than another course of action, B, pursue A over B*. A deliberation mechanism such as this is part of the computational simulation presented in chapter 8.

All of the cases of Consequence discussed so far are IRUs that are not adjacent to their antecedent. This distributional fact might be taken to be an identifying characteristic of this type of IRU. Then, according to the model of limited attention developed here, the motivation for the IRU would be that the propositional content of the IRU must be salient, but it is in fact not currently attended to. The point of the IRU is then to make that proposition currently salient.

A potential counter-example to this view of Consequence is provided in 38. This case is different from the previously considered examples because the addressee's inference process is cued by

the use of a conditional.¹ In addition, it would be difficult to argue that the redundant clause in 38-(20) isn't currently salient since it has just been discussed:

- (38) (16) h. Alright *how does the income break down?*
 (17) d. *About two thirds to one third*
 (18) h. *The two thirds I may I assume is yours?*
 (19) d. *Right*
 (20) h. If THE TWO THIRDS IS YOURS, you can get two thirds of the taxes and the interest
 (21) d. Will there be any problem in the future, let's say if her income should increase and the percentage changes?

In terms of the relevant actions and beliefs, we have:

A = Deduct two thirds of the taxes and the interest

Q = You can deduct two thirds of the taxes and two thirds of the interest

P = Two thirds of the income is yours

G = Minimize taxes, maximize income

The inferences that are licensed are that A contributes to achieving G and P SUPPORTS Q. Q is a qualifier for when action A can be taken (Schoppers, 1988). Qualifiers are conditions for when an action is applicable but don't correspond to conditions that can be achieved by an agent.

The reason that the antecedent is explicitly restated here, even though it is presumably currently attended to, is because Harry wishes to make a stronger statement. In the previous cases the inference of the rule has not been made explicit. According to the model of beliefs developed in chapter 3, this means that the inference rule is only endorsed as a default.² In this case, Harry can give the inference rule a linguistic endorsement. This follows from the fact that the relation between P and Q is a general rule, in fact part of the tax code, and Harry wants to communicate this fact.

There is an additional processing motivation for using IRUs as warrants for adopting goals or as supports for adopting beliefs. In this type of dialogue, the discussion centers on proposed future actions and beliefs that support one course of action over another. Yet agents do not necessarily share beliefs about actions and their role in contributing to different goals, and because agents are autonomous, the acceptance of a proposed course of action is not guaranteed. Therefore, while an agent may infer that a proposition is meant to serve as a warrant for adopting a proposed course of action, whether she does in fact accept the warrant is a separate issue (Webber and Joshi, 1982). Further reasoning or interaction may be required. However if the proposition that is intended to serve as a warrant is already accepted, then no further reasoning with respect to that proposition is required. This provides an excellent

¹This kind of conditional is unusual in that it violates commonly accepted felicity conditions on conditionals, which include that the antecedent be unknown.

²I'm not sure whether these should be considered relevance implicatures, but it is clear that they are weaker than entailments.

motivation for using already accepted facts in dialogue about future action.³

5.2 Alternates Relation

There is one relation between propositions that I did not discuss in the previous section because it is not clear to me whether in fact this relation has the same status as the propositional relations discussed there. I will call this the ALTERNATES relation.

What I mean by an alternate is that the IRU affirms a particular member of a salient partially ordered set as instantiating a salient open proposition (Prince, 1986). Many cases of IRUs are reaffirmations of sentences that are alternate sentences in the context in which they appear. As an example of the alternates relation, consider the excerpt below:

- (39) (30) h. Is he self-employed?
 (31) l. Yes
 (32) h. I'm sorry, i missed that
 (33) l. Yes he is.
 (34) h. Ok. Well *why not start for him a Keogh plan?* You can't get an IRA after 70, but YOU CAN GET A KEOGH PLAN

In (34) above, the first clause presumes that *you can start a Keogh Plan*. The second statement *you can't get an IRA after 70* alternates with the final one *you can get a Keogh plan (after 70)*. The pair of *IRA*, *Keogh plan* are alternates elements of a domain and task specific set defined by the property *retirement plans*. The salient open proposition is *You can (?X, polarity variable) get a (?Y, of type pension plan)*.

In the following case, a linearly ordered set is evoked of *no charges*, *low charges*, *reasonable charges*, *excessive charges*.

- (40) (22) b. Are there ah .. I don't think *the ah brokerage charge* will be ah that excessive
 (23) h. No *they're* not excessive but THERE ARE CHARGES

Harry denies that a high member of this set holds, i.e. *excessive charges*, but affirms that the lowest member, *no charges*, does not hold.

There are a number of cases that seem to be related to argument structure and to providing support for an argument, or quite commonly, to stating why the speaker cannot provide support for an argument. Horn call these cases RHETORICAL OPPOSITION (Ward, 1990; Horn, 1991); this notion is partially characterized by a condition of ARGUMENTATIVE DISTINCTNESS:

An informationally redundant affirmation Q will be discourse acceptable if it counts as ARGUMENTATIVELY DISTINCT from P in the sense that where P counts as an argument for a conclusion R, Q represents or argues for an opposite conclusion R'.

³Cohen also argues that using IRUs helps hearers' processing. Since an IRU cannot be used as a claim, it must be used as evidence for a claim (Cohen, 1987).

It seems that this condition is too strong. It doesn't cover cases where the speaker states that he cannot provide evidence to support his claim as in example 41 below, said by a person leading another along a path in the woods:

- (41) A. *Something has been through here.*
I don't know if it was a deer or what, but SOMETHING HAS. (LW 7/8/92)

Examples like this might be considered affirmations related to an evidential ordering. In the following example, the affirmation in (26) may evoke such an evidential ordering. The evidence would support the truth of Harry's assertion that *there's a line for it*.

- (42) (26) h. well you'll have to list that as interest earned
and then you can knock that amount off somewhere on schedule b,
there's a line for it,
i don't remember what the line number is,
BUT IT'S THERE

Harry's statement that he doesn't remember what the line number is fails to provide support for the fact that there is a line number, but it doesn't strictly argue against the existence of the line. Rather the use of the term *the line number* seems to presuppose the existence of said line. The ordering, if there is one, evoked here, could also be characterized as an entailment ordering since remembering what the line number is entails the existence of a line.

Many of these IRUs seem to be a reflex of a reasoning process that is comparing different states of belief and different courses of action. A number of hypotheses about the function of IRUs characterized by the alternates relation are possible:

- Reinforcement: the speaker may wish to reinforce a scalar implicature of P, namely Q.
- Set Construction: The speaker may wish to help the hearer actually construct the set that at scalar implicature could be based on, by putting set members in syntactic positions that are constrained to be related by a poset relation (Prince, 1986; Ward and Prince, 1991).
- Block Inferences: The speaker may intend to block inferences that he believes the hearer may make, such as generalization inferences (Joshi et al., 1984).
- Discourse Salience: They may function at the attentional level. By affirming P, a speaker leaves P rather than Q as the most salient proposition.
- Deliberation Reflex: these are essentially a speaker-provided protocol of a reasoning process about arguments and counter-arguments for beliefs or intentions.

The structure that is most commonly seen could result from a combination of these facts. For example a combination of deliberation and discourse salience could motivate example 42. This also seems to be the case in example 43 asserted by a passenger in a vehicle in response to the driver's comment that the heavy traffic was unexpected:

- (43) *There's something on fire* up there.
 I can't see *what's on fire*,
 but SOMETHING IS. (LW 6/12/92)

In this case, the second clause presupposes that *something is on fire* but the speaker is saying that she can't provide evidence to support her assertion. Perhaps she thinks that the hearer expects that such support should be provided. Then she returns to the main point that she intends to be discourse salient, by reaffirming her assertion (Schiffrin, 1982).

All the propositional relations that I discuss in the previous section are parasitic on the belief/intention structure, yet it isn't clear whether the alternates relation relates beliefs and intentions or whether it is a meta-level relation comparing beliefs to other beliefs and possible intentions to potential other courses of action.

Since I haven't yet determined the function of the alternates relation, I cannot include examples like these as part of the computational implementation described in chapter 8. However, agents will reason about alternative courses of action and this may capture some of the cases of what I have called alternates here.

5.3 Summary

This section discussed the Consequence function of IRUs. I suggested that agents may make inferences explicit to provide public evidence that these inferences are licensed in the context. This can demonstrate that the agent made this inference or can ensure that another agent made the desired inference.

The propositional relations of CONTRIBUTE, SUPPORT and WARRANT are used to relate the beliefs and intentions that are discussed in the dialogue, and specifies the way in which these are combined together to form a collaborative plan. An agent may produce an IRU so that one of these relations will be inferred to hold between the propositional content of that IRU and another proposition that is locally salient.

The next chapter discusses the final hypothesis, that some IRUs are a reflex of limited attention as agents' engage in purposeful dialogues.

Chapter 6

Attention

The final set of hypotheses are about Attention. The basic hypothesis is that IRUs may function to manipulate the locus of attention of the discourse participants by making or keeping a proposition salient. The fact that this is a communicative function follows from the limits on attention discussed in section 1.2 and the relation of attention to inference.

Arguably, every utterance functions to manipulate the locus of attention. It seems that the most recent utterance is usually most salient, but propositions that are frequently used should also be more salient. Here I discuss examples that function mainly at the attentional level in order to illustrate the attention function clearly.

Since attention is limited, agents must manage what goes into and gets removed from the set of items currently attended to as the discourse proceeds. This means that agents are motivated to coordinate with one another occasions when items can be swapped out of the current focus space. This coordination avoids misunderstandings that result from agents assuming different contexts. IRUs that function to introduce a topic or to close a topic demonstrate this coordination (Walker and Whittaker, 1990). For instance, consider this example from section 2.3.3 of making an entailment explicit. Note that I have labelled each sentence in Harry's turn at (31) separately, so that each sentence in the turn can be referred to.

- (44) (28) j. *I also have ten thousand in a C D*¹ six month C D
(29) h. oh uh hang on do you have anything else?
(30) j. that that's it. that's it.
(31a) h. well now we're talking about something slightly different.
(31b) YOU ALREADY HAVE SOME MONEY IN A C D.
(31c) have you anyone dependent on you?

In (31b), Harry produces an IRU whose antecedent is in (28). The use of *now* in (31a) provides support for the claim that (31a) begins a new discourse segment. The IRU in (31b) means that (31b) is to be part of attentional state for the following segment. The second type of attention examples are end-of-segment summaries as in 45-(24) below:

¹Certificate of Deposit

- (45) (7) c. Ok Harry, I'm have a problem that uh my - with today's economy my daughter is working -
 (8) h. I missed your name
 (9) c. Hank
 (10) h. Go ahead hank
 (11) c. as well as her uh husband. They have a child and they bring the child to us every day for babysitting. This is while she works
 (12) h. um hm
 (13) c. Now we're wondering how can we handle this to help them get a tax deduction and at the same time uh have my wife uh compensated for it uh so she can maybe claim social security payments or something
 (14) h. Well she's in a position where she can run herself as a self-employed individual in this regard. But there's one problem. you're gonna have to pay income tax on that money.
 (15) c. That's ok
 (16) h. You know what's good for them may not be good for you.
 (17) c. Well the way I feel that uh if it comes to say \$2000 a year
 (18) h. What is the situation. one child?
 (19) c. One child. yes.
 (20) h. Right. The maximum amount of credit that you will be able to get will be 400 *that they will be able to get will be 400 dollars on their tax return*
 (21) c. *400 dollars for the whole year?*
 (22) h. *Yeah it'll be 20%*
 (23) c. *um hm*
 (24) h. Now if indeed they pay the \$2000 to your wife, that's great.
 (25) c. *um hm*
 (26) h. SO WE HAVE \$400. Now as far as you are concerned that could cost you more. Remember you're gonna have 2000 worth of income. What's your tax bracket?

The end of segment function of the IRU in 45-(26) is correlated with the use of the cue word *so* with the IRU and the cue word *now* on the following utterance (Schiffrin, 1987; Grosz and Sidner, 1986). In this case Harry sums up the conclusion of the previous segment, which is that pursuing a potential course of action could result in a gain of 400 dollars. He then begins another segment of dialogue to compare this potential gain with the potential increase in taxes. The summary functions to coordinate the attention of the discourse participants before the transition to the new segment.

The following dialogue illustrates a return to a previous discourse segment through the use of an IRU. Eleanor (e) describes her current investments, and then states her question in (3).

- (46) (3) e.
 – and I was wondering – should I continue on with the certificates or
 (4) h. Well it's difficult to tell because we're so far away from any of them – but I

would suggest this – if *all of these are 6 month certificates and I presume they are*

(5) e. *yes*

(6) h. *then I would like to see you start spreading some of that money around*

(7) e. *uh hu*

(8) h. Now in addition, how old are you?

.

(discussion and advice about starting an IRA)

.

(21) e. *uh huh and*

(22) h. But as far as the certificates are concerned, I'D LIKE THEM SPREAD OUT
A LITTLE BIT - THEY'RE ALL 6 MONTH CERTIFICATES

(23) e. *yes*

(24) h. and I don't like putting all my eggs in one basket -

and I would suspect that february 25 would be a good time to put it into something
that runs for 2 and a half years.

that first one that comes due. Call me on the others as they come due

This is a common pattern: propositional material discussed in (4) to (6) is reintroduced into the conversation in (22). I claim that the purpose of the reintroduction is to reinstantiate these propositions in attentional state. One point of interest here is whether the clause *but as far as the certificates are concerned* is adequate to function in and of itself as a marker of a return to a previous context. It seems that it should be. In which case the function of the IRUs given above is not to mark the return per se, but rather to select particular propositions from that context as currently salient.

Note that the Attention functions are partially characterized by the fact that the IRU is remote from its antecedent. Unlike referential discourse entities that can be referred to later in a dialogue with definite referring expressions, it is difficult to use a reduced form to reintroduce propositional content (Webber, 1986); paraphrases are the main way that such reintroduction is done.

The following chapter discusses distributional patterns in the corpus and the way in which these provide support for the proposed hypotheses.

Chapter 7

Corpus Analysis

The initial analysis includes a qualitative analysis of just over a 1000 instances of IRUs from the HG corpus. This qualitative analysis is what has given rise to the hypotheses discussed in the previous chapters. I have begun to supplement this qualitative analysis with a distributional analysis. The point of a distributional analysis is to test whether a set of independent parameters can serve as diagnostics for the proposed communicative functions of IRUs. The preliminary analysis presented here is based on 163 IRUs from 24 dialogues (982 utterances).

Section 7.1 will discuss some of the distributional properties and the preliminary analysis. Section 7.2 will discuss which of my claims can be supported by a distributional analysis on this type of corpus and which cannot.

7.1 Distributional Properties

Consider the simple property of the **location** of the IRU with respect to its antecedent. The antecedent may be:

- within the same turn of the speaker, `SELF`
- adjacent to the IRU, in the previous turn of another speaker, `ADJACENT`
- in the last turn of the speaker, i.e. there is one intervening turn by another speaker, `LAST`
- remote from the IRU, in a turn that was prior to the last turn of the speaker, `REMOTE`

The location property alone picks out interesting semantic categories in the data. See Figure 7.1.

These early results indicate that:

1. Repetitions are most likely to be adjacent, (42 out of 61)

Location	Repetitions	Paraphrases	Logical Inferences	Non-Logical Inferences
SELF	0	2	2	7
LAST	8	15	2	5
ADJACENT	42	15	22	8
REMOTE	11	28	8	1
TOTAL	61	60	34	21

Figure 7.1: The distribution of various types of IRUs with respect to their location

2. Paraphrases are more likely to be remote, (28 out of 60)
3. Logical inferences are often adjacent, (22 out of 34)
4. Lasts are most often paraphrases, (15 out of 25)

The fact that repetitions are most often adjacent to their antecedents weakly supports the hypothesized Attitude function, since utterances related to Attitude must occur soon after their antecedents. Repetitions may also be more likely to be adjacent because memory for verbatim content is lost very quickly (Anderson, 1974). In addition, 35 out of the 42 adjacent repetitions are elliptical and only realize the information focus of their antecedent.

An IRU remote from its antecedent cannot have Attitude as its communicative function. The fact that Paraphrases are likely to be remote provides weak support for an analysis in which paraphrases function to manipulate Attention. However the fact that these are paraphrases rather than repetitions may only be a reflex of the fact that memory for verbatim content is lost very quickly.

The fact that many of the logical inferences are adjacent to an antecedent supports the claim that inferences must be made on propositions that are currently salient, as well as the claim that inferences can be used to demonstrate understanding.

Finally, the fact that Lasts are most often paraphrases demonstrates the function of Lasts as a recovery strategy when the speaker has reason to believe that she did not succeed in communicating what she wished to communicate. A natural recovery strategy in this context is to simply try again. Apparently speakers believe that paraphrasing an utterance increases the likelihood of success. I don't attribute this fact to lack of salience since it would be surprising if an utterance currently under discussion was not currently salient.

Figure 7.2 adds another parameter to those in figure 7.1, namely whether Harry or the Caller produced the IRU (H/C). The speaker parameter tests whether Harry produces IRUs in completely different situations than the caller does.

	Repetitions H/C	Paraphrases H/ C	Logical Inferences H/C	Non-Logical Inferences H/C
SELF	0/0	1/1	1/1	5/2
LAST	7/1	13/2	1/1	5/0
ADJACENT	12/30	14/1	18/4	4/4
REMOTE	7/4	12/5	8/0	1/0
TOTAL	26/35	51/9	28/6	15/6

Figure 7.2: The distribution of various types of IRUs with respect to their location and speaker

The fact that Harry produces most of the paraphrases and is the one making inferences explicit is consistent with Harry trying to support inferences he wishes the caller to make and manipulating the caller's attention. Since he is the one with most of the domain knowledge, he is proposing and supporting the details of the collaborative plan. Figure 7.2 also shows that the caller produces a disproportionate number of the adjacent repetitions.

While a common intuition is that adjacent repetitions function as questions, an examination of boundary tones on adjacent repetitions shows that the majority are neither syntactically nor intonationally marked as questions. See Figure 7.3.

	Harry	Caller
Rising	8	5
Falling or Level	4	25

Figure 7.3: Boundary Tones on Adjacent Repetitions

Other parameters not discussed here include whether the IRU stands in a propositional relation to another proposition, and whether it is said with narrow or broad focus. The distribution of cue words with IRUs, and how many antecedents an IRU has, partially identifies IRUs that function on the Attentional level. If an IRU has more than one antecedent, and is remote from its antecedent, then it is likely to be an end of segment summary. I have not yet tested whether IRUs that function at the Attentional level reliably co-occur with discourse phenomena that can be used as indirect indicators of segment boundaries such as cue words anaphora distribution, and topline variation (Sidner, 1979; Walker and Whittaker, 1990).

I intend to expand this distributional analysis as part of the proposed work. The most difficult part of the analysis is finding and marking IRUs and their antecedents in the transcripts since there are no automatic procedures for finding semantically related material. This has already been completed for approximately 1000 IRUs in the HG corpus, of which only 163 were discussed here.

7.2 Claims that are difficult to demonstrate with a distributional analysis

Using the HG corpus to support all my claims would prove to be difficult. For certain hypotheses, there are two problems with the corpus:

- The only possible measure of performance is the length of the dialogue and there is only one dialogue for each problem configuration. This means that it is difficult to substantiate any claims concerning increased problem solving efficiency that may be provided by IRUs since there is no comparison set.
- It is necessary to adopt a limited definition of inference in characterizing the class of IRUs examined here because it is not possible to verify what domain knowledge is shared. This limits cases of making inferences explicit to those based on information in the dialogue and linguistically motivated examples, but it would be desirable to have more tokens of this type.

Thus it is difficult to imagine tests to support the claims of IRUs functioning to manage resources such as limited attention and inference and to achieve understanding and acceptance, by resource-limited agents. This is because I am making claims about the mental states of the participants in a dialogue and it is difficult to imagine what kind of behavior could, for example, directly support the idea of different endorsements for beliefs. In addition, it is unclear how one would measure the benefit of using a particular strategy. For instance a conservative strategy should reduce the potential for misunderstanding. However, how could one calculate whether this produced fewer misunderstandings than there would have been otherwise? My conclusion is that an implementation of an explicit processing model as proposed in the next section is the only way to validate my processing claims.

Chapter 8

Computational Simulation

My claims are that INFORMATIONALLY REDUNDANT utterances, IRUs, are related to processing limitations and that they demonstrate how agents engaged in problem-solving dialogues cope with resource-bounds. I focus on a particular instance of dialogue: mixed-initiative, task-oriented dialogues between autonomous agents. The result of the dialogue is a collaborative plan, i.e. a constellation of mutual beliefs and intentions that will satisfy the agents' mutual goal (Pollack, 1990; Grosz and Sidner, 1990).

In order to demonstrate the utility of the proposed communicative functions of IRUs, my plan is to conduct a computational experiment. This experiment consists of simulated dialogues between autonomous agents with variable dialogue strategies and processing capabilities. In particular, the different dialogue strategies are primarily composed of varying whether agents produce IRUs in certain dialogue situations. The goal is to test the utility of IRUs in reducing the processing that agents perform in dialogue. This simulation will attempt to validate the putative IRU functions of Attitude, Consequence and Attention, also described in previous chapters.

Attitude (chapter 4) is what I have dubbed the communicative function achieved when agents infer whether a proposition representing a belief or an intention is mutually understood and mutually accepted. The distributional analysis of IRUs suggests that this process works over adjacent utterances, or over a limited number of utterances. One proposed reason for local management of this process is to ensure that agents commit to certain mutual beliefs and intentions, so that further reasoning is reduced (Bratman et al., 1988).

Consequence (chapter 5) refers to utterances in which inferences are made explicit as well as to utterances that license the inference of a propositional relation between two propositions. The inference of propositional relations, such as CONTRIBUTE, WARRANT and SUPPORT, seems to depend on two propositions being salient at the same time.¹ Thus if one proposition isn't salient, it may be restated to make it salient. An IRU can save processing for one agent in a dialogue, by avoiding situations in which both agents have to make the same inference by the

¹This is compatible with theories of how rhetorical relations are inferred, such as RST. It is also compatible with a complexity proof by Vardi that shows that it is the binary rules of inference, allowing the combination of two separate facts from the database, that increases the complexity of epistemic reasoning (Vardi, 1989).

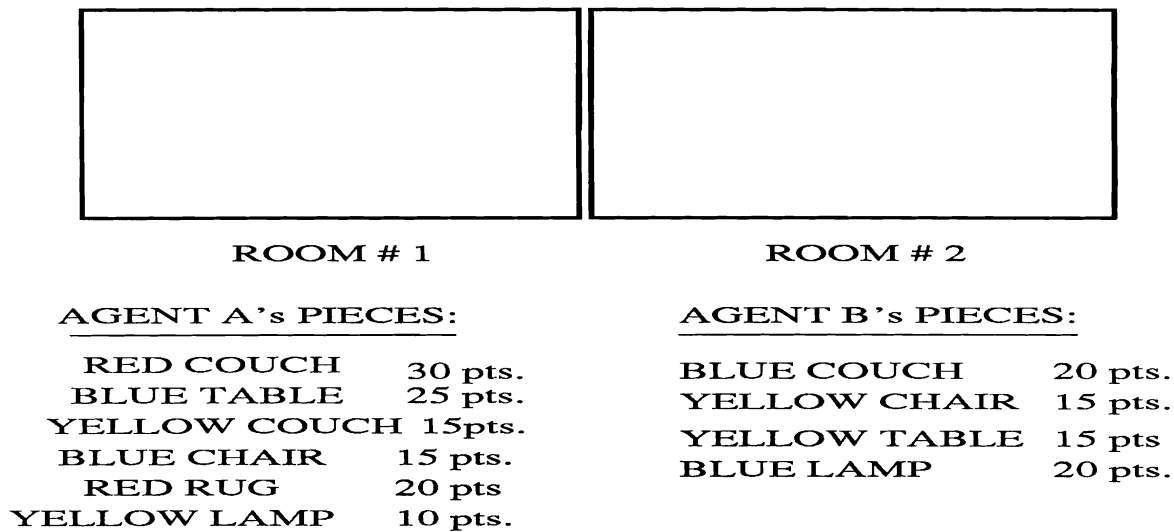


Figure 8.1: A Sample Initial State for the Design-World Task

afraid I can't help you. Good bye. Furthermore, in the HG the corpus there is only one instance of any problem and only one dialogue about it. This means that it is impossible to compare a case where one of the conversants produces IRUs with cases where they don't.

I propose to conduct the simulation in the Design-World artificial domain using a variation of a cooperative task that has been the basis for experiments on distributed cooperative work(Whittaker et al., 1992). The world consists of the floor plan for a house. The task involves two agents with a mutual goal to arrange a number of pieces of furniture in the rooms on the floor plan. Each agent has different pieces of furniture and there are more pieces available than needed to accomplish the task. Each piece of furniture has an associated point value. The result of the agents engaging in a dialogue is a collaborative plan for a design layout for the rooms on the floor plan. The agents attempt to maximize the points achieved by the collaborative plan, while doing the task within the given constraints.

Despite the disadvantages of not being a real domain, Design-World has a number of advantages. The task is inherently mixed-initiative and exhibits properties of real distributed design tasks(Suchman, 1985; Bly, 1988; Whittaker et al., 1992). It is simple; there is a limited amount of domain knowledge and only a limited number of actions to reason about. Yet it provides a natural set of inferences in terms of calculations to maximize utility and the subgoals remaining to complete the design for a room. Performance is easily measured by the points associated with the final design as compared with the costs to achieve these points. Costs will be based on: (1) number of utterances in the dialogue; (2) the amount of means-end reasoning required; and (3) the number of steps involved in searching the representation of current beliefs and intentions. This makes it possible to explore in a principled manner the trade-offs associated with different dialogue strategies.

In order to have some task structure, the basic Design-World task consists of a pair of agents achieving a design for a floor plan with two rooms. Furniture items are of 5 types: couch, table, chair, lamp and rug. Each furniture item has a color and point value. A design for a

room consists of any four pieces from these types. The points associated with a furniture item supports the calculation of utility of including that item in the design plan. This is what the basic performance measure is based on. Figure 8.1 shows the initial state for the dialogue.

Each agent starts with private beliefs as to what pieces of furniture they have, what colors these pieces are and how many points they are worth. The associated domain state predicates are:

- (Has ?Agent ?Furniture ?Time)
- (Points ?Furniture ?Value)
- (In ?Room ?Furniture ?Time)
- (Design ?Room)

In other words, what is shown in figure 8.1 as Agent A's piece, e.g. a red couch worth 30 points, would be represented by two assertions: (Has A Red-Couch Now), (Points Red-Couch 30). The associated domain actions are:

- (Put ?Agent ?Furniture ?Room ?Time)
- (Remove ?Agent ?Furniture ?Room ?Time)

Some states are achievable; they can be converted to actions by an Achieve predicate, e.g. (Achieve ?Agent (In Room1 Red-Couch)) and (Achieve ?Agent (Design ?Room)). Agents will be taken from a set of two, Ann and Bob, as well as Ann and Bob as a pair, A&B. The ?Time variable for each action and state is either NOW or LATER, and so is only able to distinguish future actions from current actions.

In the Design-World simulation, the agents will share beliefs about what it means to have a COLLABORATIVE PLAN to achieve a mutual goal. A COLLABORATIVE PLAN is defined as:

- (COLLABORATIVE-PLAN A&B (Achieve A&B Goal)) \iff
 - 1. (MB A&B (Intend A&B, $\alpha_i \wedge \dots \alpha_n$))
 - 2. (MB A&B (Contribute α_i (Achieve A&B Goal))
 - 3. (MB A&B (Max-Utility $\alpha_i \wedge \dots \alpha_n$) (Achieve A&B Goal))

A sample subgoal of the whole task would be (Design Room-1), and this would instantiate the definition of a collaborative plan given above with domain knowledge that a room design consists of different combinations of 4 pieces of furniture. The agents have to mutually believe that: (1) they intend to put each piece of furniture into a room; (2) these actions of putting furniture into a room contribute to their mutual goal; and (3) the actions they have chosen have the maximum utility they are able to achieve given the constraints of the task, their

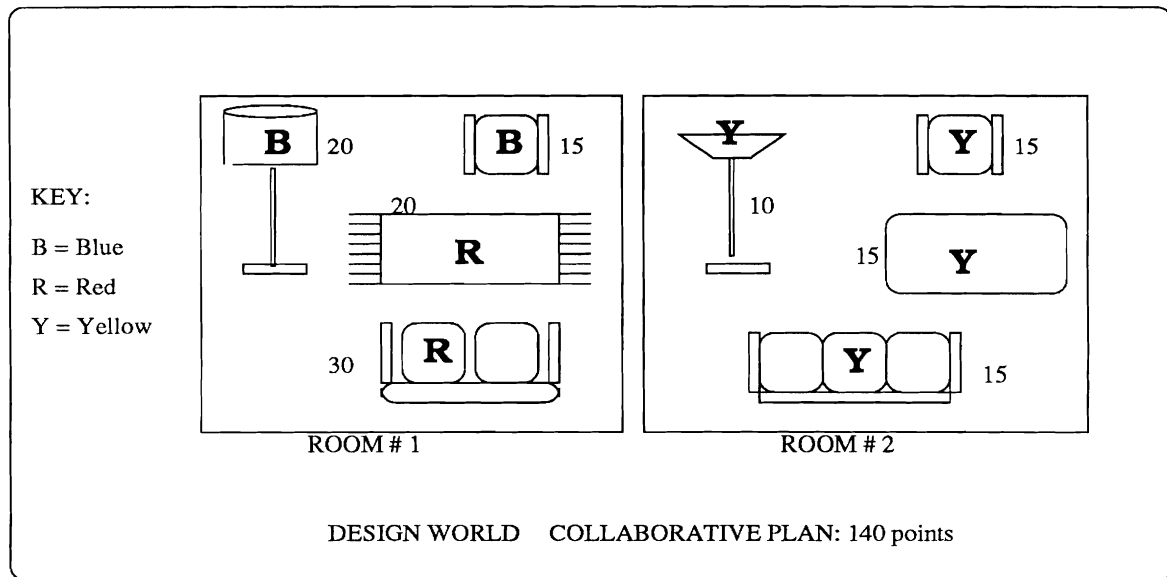


Figure 8.2: Potential Final State for Design-World Task: Represents the Collaborative Plan Achieved by the Dialogue

cognitive limitations, and the externally imposed time limits on reasoning about how to do the task (means-end reasoning). The account of mutual beliefs was described in chapter 3. Details of the implementation will be given in section 8.3.

Under the task definition, a predicate to produce subgoals, *MATCHED PAIR*, is defined as an optional component of the (Design ?Room) goal:

If the design includes a *MATCHED PAIR*, then the points of each piece in the pair are doubled.

A restriction on this rule is that an item may only participate in one *MATCHED PAIR* subgoal. The color predicate provides the additional parameter that supports these subgoals. Subgoals are necessary in order to provide some added complexity to the discourse structure. For instance, a matched pair subgoal can motivate temporarily suspending a discussion about the design for one room, in order to try to achieve a matched pair for the other room. This kind of complexity helps to motivate the Attention type (context shifting) IRUs. In addition, this complexity means that the calculation of utility is not simply the addition of the values of the pieces involved. An individual piece may have a high value but there may be another piece that is worth more to the final plan because it is able to combine as a matched pair with another piece.²

Initially an agent has a number of pieces of furniture. A *COLLABORATIVE-PLAN* is achieved by a cycle of: (1) individual agents perform means-end reasoning about options in the domain;

²This added utility from combinations may also provide an approximation of Galliers' claims that the coherence of a belief set is more than the summed coherence of the individual beliefs (Galliers, 1990; Harman, 1986).

(2) individual agents deliberate about which options are preferable; (3) then agents make proposals to other agents, based on the options identified in a reasoning cycle, about actions that contribute to the satisfaction of their goals; (4) then these proposals are accepted or rejected by the other agent. For instance Ann may conduct means-end reasoning for a set period of time, deliberate on the options produced by this reasoning, select the red couch as the best option to pursue, and then propose to Bob that they put the red couch in room-1. Bob considers this proposal and compares it with other options he knows of. Then Bob may accept the proposal, reject the proposal, or leave the proposal on the table by requesting additional information. The communicative strategies associated with making and accepting or rejecting these proposals and with acquiring and providing information on which to make decisions will be discussed in section 8.5. Figure 8.2 shows a potential final collaborative plan for the task as a result of the dialogue, based on the initial state given in figure 8.1.

Design-World is complex enough to exhibit the critical properties of the HG corpus. Dialogues are necessarily task-oriented, but there is no rigid task-structure since either room can be designed first and any design goal can be achieved before any other. This increases the variability in the dialogues and makes them complex enough to demonstrate the function of IRUs. Finally, the fact that it is a visual domain means that it is easily simulable. Even though I don't plan on having the agents execute the plan, I plan to display a visual representation of the evolving design as the dialogue proceeds.³

The expectation is that there will be trade-offs between performance and the cost measures, such as number of utterances required to achieve a solution, according to variations in communication strategies, agent's cognitive capabilities, and uncertainty in communication. The expected trade-offs according to communicative strategies and other parameters will be discussed further in section 8.5.

8.2 IRMA architecture for Resource Bounded Reasoning

The agent architecture assumed here is patterned on the IRMA architecture for resource-bounded agents (Bratman et al., 1988; Pollack and Ringuette, 1990). This architecture forms the basis for the Tileworld simulation environment, and seems appropriate as a framework for the simulation proposed here. The basic components of the IRMA architecture are:

- Beliefs: a database of an agent's beliefs
- Intentions: a database of an agent's intentions
- Plan Library: what an agent knows about plans as recipes that can achieve an agent's goals.

³There are a number of variations that could be constructed in this domain if it were necessary or desirable. For instance, the complexity of the task is easily varied by manipulating constraints or restrictions on the sequencing of actions, e.g. the couch must go in the room first. Extra constraints or conflicts could be added by giving agents preferences for putting a particular piece of furniture in a particular room. If the simulation included both a planning and an execution phase, these could be interleaved or done in two phases, and it would be possible to explore which parts of the collaborative plan were determined at planning time and which at execution time.

- Means-end reasoner: considers how to fill in existing partial plans, proposing options that serve as subplans for the plans an agent has in mind. In Design-World as in Tileworld, the time to perform means-end reasoning will be subject to an externally imposed time limitation as an experimental parameter.
- Opportunity analyzer: proposes options in response to perceived changes in the environment. (Currently in Design-World, changes in the environment will only arise as a result of communication.)
- Filtering Mechanism: consists of two parts:
 - Compatibility Filter: checks options for compatibility with the agent's existing plans. Options deemed compatible are passed along to the deliberation process. Incompatible options are passed to the Filter override mechanism.
 - Filter override: May pass an incompatible option, along with the intention that it is incompatible with, to the Deliberation process for reconsideration. Conditions on when an override may occur are encoded in this component.
- Deliberation Process: decides which of a set of options to pursue. Whenever multiple options are passed from the filtering mechanism to the deliberator they are weighed against each other. In Design-World, the deliberation strategy always involves a calculation of the point value associated with an option.

Figure 8.3 shows the components of the IRMA architecture as they apply to Design-World. In the Design-World simulation environment, I incorporate aspects of the IRMA architecture such as separating means-end reasoning from deliberation, and filtering new options to allow current intentions to play a role in reducing further deliberation. However, I wish to incorporate concepts of limited attention which will act to constrain retrieval of current beliefs and intentions and which will interact with the means-end reasoner.

In the Tileworld simulation using the IRMA architecture, both the agent and the environment were highly parameterized. The agent was parameterized with respect to the filtering and deliberation mechanisms that she used. The degree to which the simulation environment was dynamic and unpredictable was also parameterized. The Tileworld simulation explored the hypothesis that the effectiveness of an agent's meta-level reasoning strategy is dependent upon the characteristics of the environment in which the agent is situated. In Design-World, the environment consists of both the domain-based task and the other agent. The domain portion of the environment is more stable than in Tileworld, but there is uncertainty in communication. The degree of uncertainty in communication can be parameterized in Design-World just as the degree of uncertainty in the domain was parameterized in Tileworld. Communication also serves to increase the dynamic nature of the environment. It is impossible for an agent to predict what better options another agent might have available to contribute to their collaborative plan.

In the Design-World task, the potential intended acts, ie. options, are domain level intended acts such as putting a particular furniture piece into a particular room. These acts however correspond to intended proposals in some cases, but may correspond to proposals that have

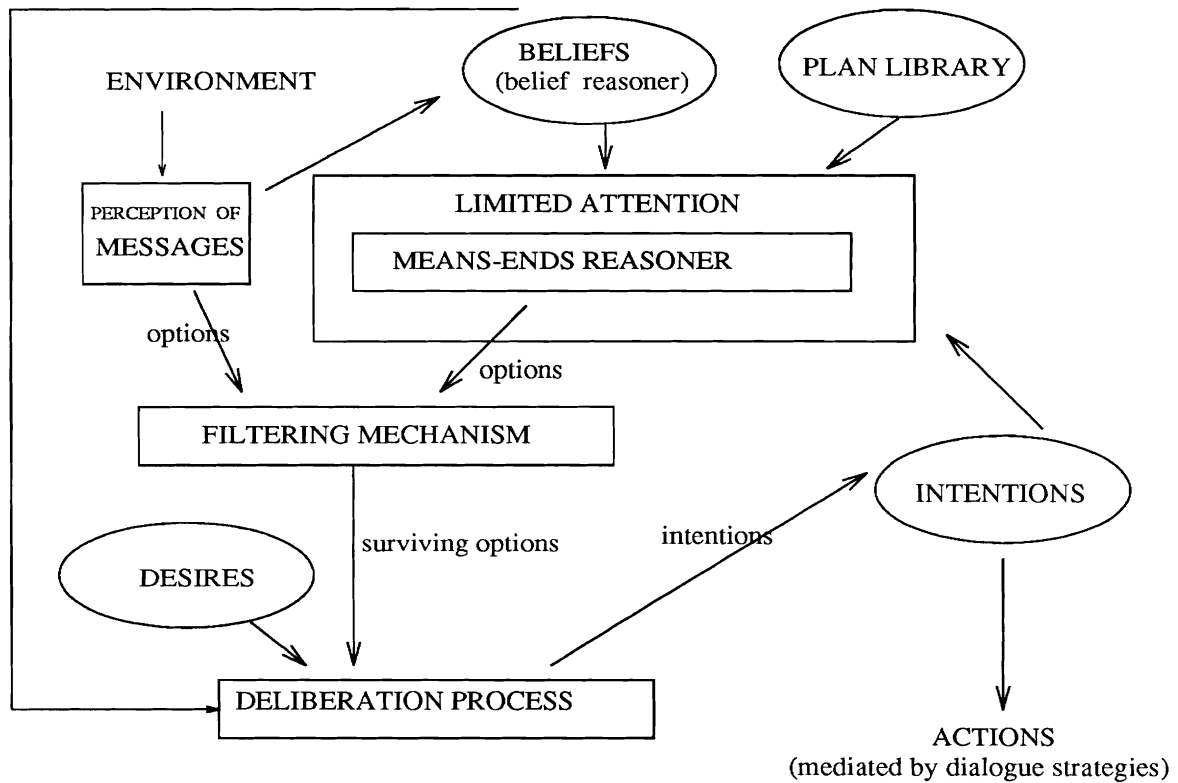


Figure 8.3: Design-World IRMA Agent Architecture for Resource-Bounded Agents

already been made and accepted. Because the Design-World simulation focuses on multi-agent communication, beliefs will be annotated as to whether they are thought to be mutually believed and agent's intentions are annotated as to whether the agent believes they are mutually intended. The structure of these beliefs and intentions will be discussed further in sections 8.3 and 8.4.

The Filtering Mechanism must be carefully designed. One key aspect is that the Compatibility Filter must be computationally tractable and it must not duplicate the functionality of the deliberation mechanism. Furthermore, in a two agent system, options that arise due to another agent's messages must at least be responded to, even if they do not make it to the deliberator. Another key aspect is that the Filter Override must be carefully designed so that agents' existing intentions act to constrain further reasoning and endless reconsideration of intentions, while at other times allowing intentions to be reconsidered. The Filter Override can encode any conditions that are deemed desirable to let incompatible options through. Tileworld's simplest Filter Override condition compares the score of a proposed option to that of the option that is currently being pursued. If the difference between them equals or exceeds some threshold value v , then the new option passes the filter. The threshold value is set as a Tileworld parameter.

The Filter Override mechanism in Design-World must treat mutual beliefs and mutual intentions differently than private ones because they play a role in the collaborative plan that is

being constructed. Agents must communicate with other agents about potential revisions to these beliefs and intentions. A mutually intended proposal must be retracted and agreement about the replacement proposal must be achieved, meaning that the cost associated with the communication required to establish agreement in the first place is wasted. In the first simulation experiments, the filter-override for both agents will be designed so that actions incompatible with mutually accepted proposals never make it through the filtering mechanism. This can be implemented with a filter threshold of positive infinity for mutually agreed proposals.

The next section briefly discusses the belief representation system that will underly the representation of mutual beliefs.

8.3 Belief representation system

This account requires a belief representation system that can keep track of beliefs and their underlying assumptions, as well as the endorsements on these assumptions. Galliers has developed a theory and a system for autonomous belief revision called ABR that has these properties (Galliers, 1990; Galliers, 1991b; Galliers, 1991a; Cawsey et al., 1992). The beliefs model developed here is a version of ABR. ABR incorporates a specialized ATMS developed for dialogues between autonomous agents. Each belief has an associated endorsement. I use the endorsement types of `HYPOTHESIS`, `DEFAULT`, `ENTAILMENT`, `LINGUISTIC`, and `ABSOLUTE` as outlined in chapter 3. This system also implements the Principle of Positive Undermining discussed in section 4.1.

However ABR requires a few modifications to be used as part of the account presented here. First, ABR doesn't reason about mutual beliefs. Gallier's concern has been an individual agent's reasoning as to whether to revise her own beliefs by either accepting or rejecting some incoming information. Second, in Gallier's system, agents only communicate to resolve conflicts in their own beliefs, not to achieve mutual beliefs. Third, agents in ABR are logically omniscient.

In my account, each agent has a version of mutual beliefs that depends on what has been made public in the current conversation, but may include propositions licensed as inferable. These representations will be built on a modified version of Gallier's ABR. Each dialogue act and inference is tracked by ABR with respect to its underlying assumptions. The extensions to ABR explored here are limiting inference, the notion of limited attention, the exploration of the weakest link rule, and the incorporation of dialogue principles to reason about mutual beliefs such as the collaborative principle and conversational defaults discussed in section 4.1.

8.4 Discourse Actions

There are three aspects of communicative actions that must be specified: (1) the utterance level intentions that an agent can achieve, (2) the propositional content of these utterance level intentions, and (3) the way in which these intentions are reasoned about. I will discuss the first

two of these in the remainder of this section. The way in which these intentions are reasoned about is encapsulated in agents' communicative strategies. These strategies will be discussed in section 8.5.

For the purpose of this simulation, agents communicate through an artificial language by sending messages to one another. I implement agents within the same process and use a controller to switch control between them.⁴

I use 5 utterance level intentions that may be specific to this task or this type of dialogue (See also (Carletta, 1992; Sidner, 1992)). These are Propose, Accept, Reject, Ask and Say. It would be possible in this domain to leave these intentions implicit and only communicate propositional content. Then hearers' would need to infer whether an utterance counts as one of these actions. If this were done, additional inference procedures would be needed to distinguish implicit acceptance from rejection. Since these inferences are not the focus of this thesis, I make these intentions explicit. Each communicative act of an agent fits the schema given below:

(?Utterance-Intention ?Agent1 ?Agent2 (?Pred ?Vars))*

The basic schema consists of a proposal to do a particular action. The predicates in this domain are: (1) the domain action predicates of Put and Remove, which correspond to proposals that the speaker is making; (2) the state predicates Points, Has, In, Design. For instance agent A might say:

(Propose A B
(Put A&B Red-Couch Room-1 Later))

Agent B can either accept, reject or ask for more information about this proposal. For example, B could reply with an acceptance in the form of a repetition:

(Accept B A
(Put A&B Red-Couch Room-1 Later))

B can also ask for more information:

(Ask B A
(Points Red-Couch ?))

As a simplifying assumption, interaction is limited such that B will always ask for more information if B doesn't have enough information to make a comparison between the options that B knows about and the one that A has proposed. However, if B has no options available to satisfy the same goal, B can accept A's proposal without knowing its exact value.

B can also reject the proposal:

⁴This is a common way of simulating individual processes and has been used in the Tileworld simulation, in Power's robot world and in Carletta's JAM system (Pollack and Ringuette, 1990; Power, 1974; Carletta, 1992). There seems to be no need for the added complexity of two separate processes.

(Reject B A
 (Put A&B Blue-Couch Room-1 Later))

The predicate clause in a rejection is a counter-proposal that encodes the reason that B rejects the proposal that A made. In this case, B believes that putting the Blue-Couch in Room-1 is a better option. The production of a counter-proposal is a common, but not the only way that rejections are done in human-human dialogues. In the proposed simulation, rejections always include a counter-proposal. This is because, in the Design-World task, the decision to reject a proposal is based on comparing the proposal with other options an agent knows about. Another option must be of greater utility to support a rejection, and since rejections are rarely IRUs, I don't intend to implement different ways of communicating a rejection. In addition, designing rejections to work in this way allows for a change of initiative with each rejection, which is very common in human-human problem solving dialogues (Walker and Whittaker, 1990).

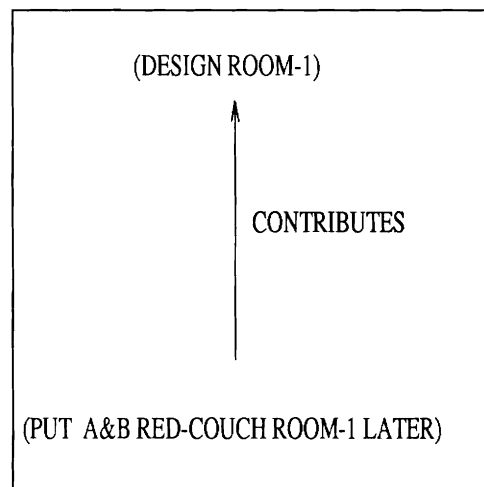


Figure 8.4: Partial Plan for Achieving the Design of Room 1

The utterance intention of Say is used to communicate facts about the current state, and to reply to questions. For instance if B asks A:

(Ask B A
 (Points Red-Couch ?))

A can reply with:

(Say A B
 (Points Red-Couch 30))

As messages are sent and received during a dialogue, agents construct a corresponding belief/intention structure. The propositional relations of CONTRIBUTE, SUPPORT and WARRANT

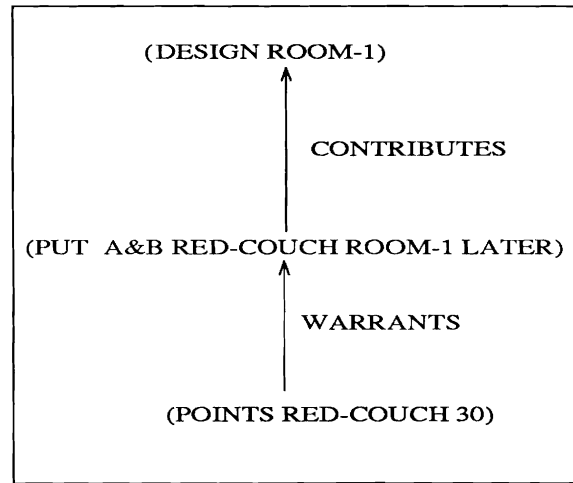


Figure 8.5: Partial Plan for Achieving the Design of Room 1 with a Warrants relation

are the glue that holds this structure together. These can be viewed as labels connecting communicated propositions. The goal that a communicated proposition is related to can be inferred from the ?Room argument of a proposal. Then if the utterance intention is a proposal, the propositional relation is inferred to be a **CONTRIBUTE** relation. For example, if A says:

(Propose A B
(Put A&B Red-Couch Room-1 Later))

B can infer that the proposed action **CONTRIBUTES** to the (Achieve A&B (Design Room-1)) goal. If B accepts this proposal, then A and B mutually believe that the belief/intention graph shown in figure 8.4 is part of their collaborative plan.

Similarly, any communicated proposition involving the Points predicate must relate that proposition to a goal via the **WARRANTS** relation. For example, if it were mutually believed that (Points Red-Couch 30), then the belief/intention structure might be as depicted in figure 8.5.

Section 5.1 discussed the fact that propositional relations in real dialogues are sometimes made explicit and sometimes left to inference. In the communication language proposed here, I leave them implicit. In the general case, the recognition of these relations is a plan inference problem and is beyond the scope of this thesis (Sidner and Israel, 1981; Sidner, 1983; Sidner, 1985). However, the situation is restricted enough in the Design-World task that these relations can generally be easily inferred. As discussed above, the goal can usually be inferred from the form of the proposal. The only situation in which this is not the case is when an action can simultaneously function to contribute to multiple goals. For example, if the action also could contribute to a matched-pair goal, and this has not been made explicit, the hearer could fail to make this inference. The hearer will not however be unable to make sense of the proposal, because it is trivial to infer one of the goals that the proposed action contributes to. This then looks like a case of whether Consequence is made explicit or left to inference. This will be discussed further in section 8.5.3 and 8.5.2.

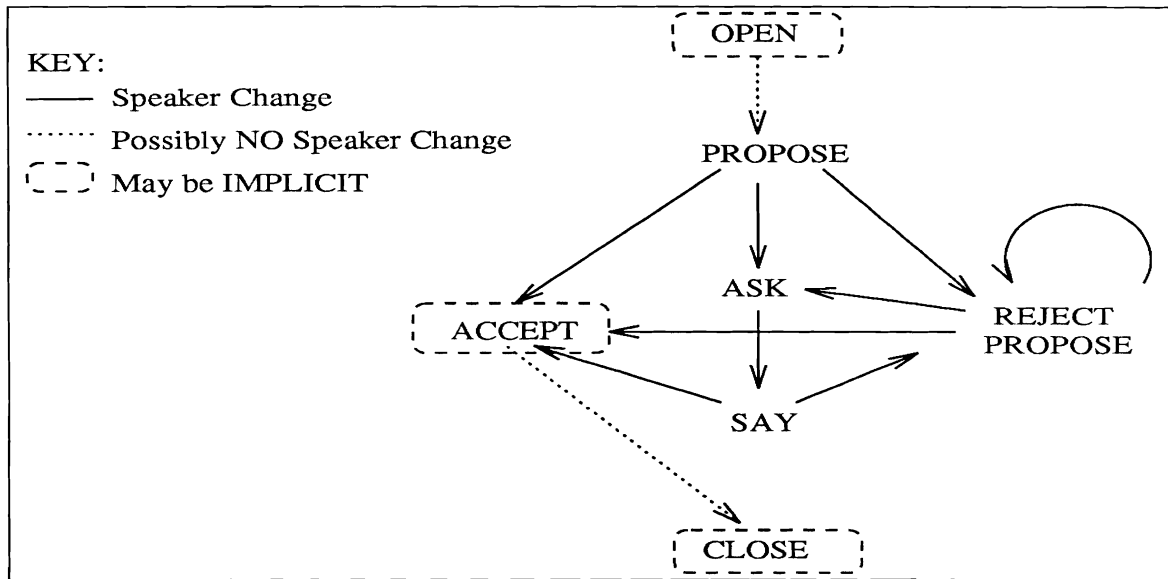


Figure 8.6: Dialogue Actions for the Design-World Task:Version 1

The dialogue actions and the potential responses are shown in figure 8.6.⁵ The open and close dialogue actions shown in figure 8.6 are not always explicitly realized. They will be discussed in more detail in section 8.5.3. As the figure shows, proposals can be followed by acceptances, asks, or rejections. Rejections always include counter-proposals. In addition the form of proposals, and acceptances will depend on the dialogue strategies discussed throughout section 8.5. The potential dialogue strategies also must be extended to allow a proposal under consideration to be suspended while another proposal is considered. This will be discussed further in section 8.5.3.

8.5 Agent Communication Strategies

An agent's communication strategy determines which discourse actions an agent has available, and in particular whether they produce IRUs in certain discourse situations. The hypotheses about the benefits of IRUs will be tested by comparing the performance of agents that systematically produce IRUs in certain discourse situations as compared with agents that don't. Since it is necessary to separately test my hypotheses about the three separate classes of IRUs, there are strategies associated with each of Attitude, Consequence and Attention. However in order to keep the simulation manageable, the variations in strategies must be simpler than those characteristic of real dialogues. Figure 8.7 depicts schematically the dialogue actions that the strategies affect. The Attention strategies will mainly target segment openings and closings. The Consequence strategies will mainly target the way in which Proposals are made and

⁵These actions do not represent a claim that this is how humans conceptualize dialogue or that problem solving dialogues can be analyzed into structures exactly like these. These are just an approximation in order to investigate the effects of resource-bounds, which will operate independently of the range of utterance level intentions and other strategies available to an agent(Levinson, 1979; Levinson, 1981; Schegloff, 1987).

the Attitude strategies will affect the way acceptance is communicated. Of course as I noted earlier, things are not so simple in real dialogues. Both the Attitude strategies and the Consequence strategies may interact with limited attention. This interaction will be investigated in the simulation.

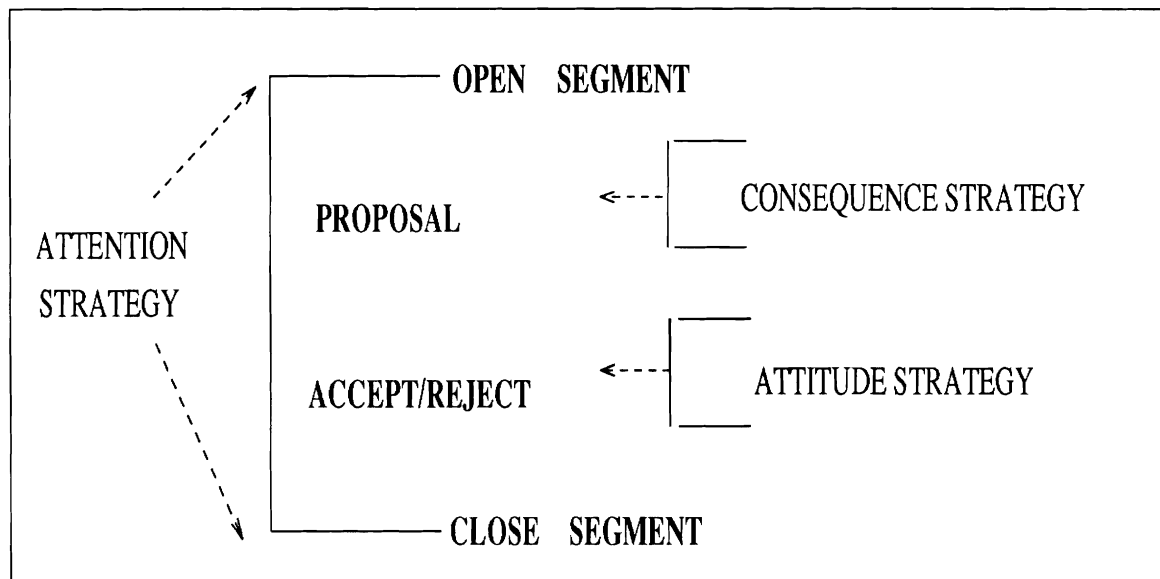


Figure 8.7: Schema for the Primary Effect of Communicative Strategies on the Dialogue Structure

Initially, two variations of explicit compared to implicit, for each of Attitude, Consequence and Attention, will be explored. These variations in strategies will be discussed in detail below. Here I just give an overview of the full scope of the proposed experiment. The prediction is that these communication strategies correspond to agents' cognitive properties and will interact with them.

Each strategy will be implemented by variations in the messages that agents send one another. These messages will encode the communicative functions for IRUs discussed above, but will not vary as to the different forms that realize these functions. This variation in form is a topic for future work.

In order to see the effect of the Attitude strategies, it is necessary to have some variation in the communication environment to demonstrate the benefits of conservative communication strategies in uncertain environments. Uncertainty in communication will be implemented so that the degree of uncertainty (noise) can vary from cases where communication is certain to cases where almost every utterance has the potential for an error either in transmission or in interpretation.

In order to see the effect of the Consequence and Attention strategies, it is necessary to vary the amount of time that agents can perform means-end reasoning and to track the time required to search current beliefs and intentions. The number of steps involved in searching current beliefs is related to the Attention strategy, and the number of inference steps conducted is related to

the Consequence strategy. No matter how the current beliefs and intentions are structured, the hypothesis is that an explicit Attention strategy means that virtually no search is required, whereas some search will be required otherwise.

However it is clear that the time to search current beliefs depends on the way they are structured. In terms of a cognitively accurate model, this is an open research problem. The question is whether some approximation to the structure of memory can be adopted that will allow a valid measurement of the effect of the Attention strategies. One potential simplifying assumption is that current beliefs are structured in chronologically linear order based on recency. Search would then take place in reverse chronological order. However research on attention and memory in psychology indicates that frequency also has an effect on retrieval. It seems that some way of including frequency effects in the memory/attention model used here must be determined.⁶ I am currently investigating several models of attention and memory to determine whether there is a simple, cognitively plausible model that could be used in the simulation (Landauer, 1975; Just and Carpenter, 1992).

The total space of the simulation is rather large, even with the limitations proposed of only two variations for each of the communicative functions of Attitude, Consequence and Attention. These combinations would give 8 different agent type combinations in terms of communicative strategies. Adding in two variations for noise in communication and two variations for the time allowed for means-ends reasoning would give 32 potential experimental set-ups. No variations are required for Attention since time to search current beliefs and intentions will just be tallied. Thus the total number of potential dialogue simulation types is 32 if: (1) the task is constant; (2) the information that each agent has at the beginning of the dialogue is constant, and (3) if agents always interact with only one kind of other agent. However, I don't intend to explore the whole space. I will focus on certain points in the space that I expect to produce interesting interactions, e.g. the interaction between an agent that has an explicit Consequence strategy and one that is a limited reasoner. Additional runs can be done if there is time to do a fuller interactional analysis. For instance, it would be of interest to change the information distribution and see how the dialogue varies.

A question that comes up repeatedly in analyzing dialogue and in producing this kind of simulation is the difference between a strategy that makes it possible for an agent to do something as compared with one that makes it easier for an agent to do something. It is often impossible to distinguish these two cases from observations of the behavior of human agents. For instance when analyzing the utterances in a dialogue, we cannot distinguish eliminating ambiguity from reducing the amount of search or inference required to unambiguously interpret an utterance. However in the simulation environment we can state whether in principle it is possible for an agent to interpret an utterance and compare this possibility with the amount of time that it takes to make the interpretation.

The communicative strategies associated with each of the three IRU classes will be discussed below. Section 8.5.1 will describe the proposed Attitude strategies, section 8.5.2 will describe

⁶A complication with a representation that includes only recency and frequency for directly realized propositions is that it is possible that retrieval of one belief would bring along and 'refresh' other beliefs and intentions that are associatively linked by propositional relations such as WARRANTS. This may be a version of a focus space; the mention of something introduced in a particular context evokes the whole context and not just the item mentioned (Grosz, 1977; Sidner, 1979).

the proposed Consequence strategies and section 8.5.3 will describe the proposed Attention strategies.

8.5.1 Attitude Strategies

Earlier I discussed 5 different dialogue actions that directly address the assumptions underlying mutual understanding (Clark and Schaefer, 1989), and which also can lead to an inference of acceptance by the application of the COLLABORATIVE PRINCIPLE:

- prompt such as *uh huh*
- repetition
- paraphrase
- making an inference explicit
- implicit acceptance, going on with next part of the plan

In order to investigate some of the trade-offs with these different ways of indicating acceptance, it is necessary to contrast implicit acceptance with one of the attitude-explicit dialogue actions. The choice of which attitude-explicit action to implement interacts with the choice about how to make communication less than certain. If communication is uncertain due to referential ambiguity, then paraphrase is a good strategy. If it is desirable to model something similar to the noise in communication in a speech recognition system, then repetition is a reasonable attitude-explicit action. I don't consider using prompts because they only demonstrate attention, and any relevant utterance can demonstrate attention. I don't consider the strategy of making an inference explicit because it interacts with Consequence. Because paraphrase requires a source of principled ambiguity, I will implement REPETITION as the attitude-explicit action.

Thus there will be two agent strategies associated with Attitude.⁷ The two attitude strategies are: (1) IMPLICIT ACCEPTANCE: the agent indicates acceptance by going ahead to the next part of the dialogue, either by closing the current segment of dialogue or by making a new proposal; (2) the agent indicates acceptance with a REPETITION of the proposal that the other agent made.

As mentioned above, the REPETITION explicit Attitude strategy interacts with a specific kind of noise in the communication environment. Noise in the transmission and interpretation of a message will be simulated with a noise-maker which will randomly pick an argument of a communicated predicate and replace it with another value of the same type. For example, a proposal to put the red chair into room-1, could be changed into a proposal to put the blue chair into room-1, to put the red couch into room-1, or to put the red-chair into room-2. This means that the errors will produce messages that are interpretable in the domain. A problem with this kind of error is that it may cause some bias toward the explicit Attitude strategy

⁷For simplicity, agents in the simulation always reject a proposal in the same way, so rejection introduces no variation.

because it will increase the likelihood of that the error is not discovered immediately. This may not be a very good approximation of the type of errors that one might actually expect, but the effect of the error and the recovery from the error should approximate real dialogue situations closely enough to test the hypothesized attitude function.

The prediction is that in noisy communication environments, the explicit attitude strategy embodied by repetition will be more efficient since it avoids potential invalid plans, or at the least costly misunderstandings and replanning. In order to demonstrate this properly, it is necessary to have a reasonable recovery mode for errors that arise from using the higher-risk strategy of implicit acceptance (Carletta, 1992).

The beliefs model proposed in chapter 3 supports recovery strategies based on examining assumptions that are weakly endorsed and therefore easily defeasible. In the case of this simulation, the choice of recovery strategy is determined by the attitude communication strategy. There is only one way in which the communication can fail, so recovery can consist of simply repeating the misunderstood utterance or the part that was misunderstood. In other words, recovery consists of addressing the weakly endorsed COMPLETE HEARING assumption.

An independent question is what happens if the misunderstanding is discovered much later in the dialogue? What happens to the intervening partial plans that have been agreed on? In human-human dialogues, this kind of misunderstanding often results in agents starting again from the beginning. As I pointed out in chapter 4, the Principle of Positive Undermining predicts that each inference made from a false belief must be explicitly retracted in order to ensure that it is gone (Harman, 1986; Galliers, 1990). For the time being, I will not require explicit retraction, but will dispose of dependent beliefs as tracked by the TMS implementing the beliefs model.

However, all the previous discussion about recovery strategies assumes that errors will in fact be detected, but this is not necessarily the case. In the case of an explicit attitude strategy, errors will always be detected immediately. In the case of an implicit attitude strategy, errors may be detected much later, or not at all. They can be detected later if the fact communicated in error is made explicit later in the dialogue via an explicit consequence strategy or attention strategy. If they are not detected at all the collaborative plan will be invalid.

8.5.1.1 Prediction

The prediction is that an agent embodying the explicit Attitude strategy will achieve higher task scores than an agent embodying the implicit Attitude strategy in communication situations that are noisy. The number of utterances produced will be tracked as a cost as well as the costs to retrieve current beliefs and intentions. I would expect that implicit Attitude strategies could result in longer dialogues in noisy situations because some of the dialogue is wasted correcting a misunderstanding that could have been corrected earlier. Alternately an implicit Attitude strategy may result in invalid collaborative plans which could either have a score of 0, or have the total score reduced by eliminating the invalid proposal.

Another prediction is that an explicit Attitude strategy may interact with the attention/working memory model since it increases the frequency of a proposition in memory. A proposition that

is repeated may be faster to retrieve. If a repeated proposition is used later on in the dialogue, the explicit Attitude strategy might result in faster retrieval times, and lower costs.

8.5.2 Consequence Strategies

As noted in section 8.4 the propositional relations of *CONTRIBUTE*, *SUPPORT* and *WARRANT* will always be left implicit and will be inferred by the hearer. I claimed in chapter 5 that one function of IRUs is to make it easier to make these types of inferences.⁸ This seems to be the case most often in the HG corpus when the IRU provides a *WARRANT* for the course of action under discussion (Cohen, 1987).⁹ Therefore the Consequence strategies are independent of these inferences which will always be left to the hearer to make.

There will be two Consequence strategies, an explicit Consequence strategy and an implicit Consequence strategy. In the implicit Consequence strategy, agents will not make inferences explicit such as the mathematical calculation to determine the utility of the current proposal. They also won't restate propositions that don't involve calculations, but that are meant to be inferred as a *WARRANT* for a proposal under discussion. In the explicit Consequence strategy, they will. In other words a proposal under the implicit consequence strategy will be:

(Propose A B
(Put A&B Red-Couch Room-1 Later))

A proposal under the explicit Consequence strategy includes information that is meant to serve as a *WARRANT*. This information is stated whether or not it is already mutually believed. For example the Points information in the proposal below:

(Propose A B
(Put A&B Red-Couch Room-1 Later))
(Say A B
(Points Red-Couch 30))

The Consequence strategies are predicted to interact with a resource-bound on agents' inference/reasoning capabilities. The limit that will be explored in the simulation is to limit agents' means-end reasoning. This is a reasonable place to put a limitation since means-end reasoning produces options that correspond to potential proposals. The structure of the domain means that some proposal can be made even with means-end reasoning severely limited. An agent only has to be able to access one predicate describing one of their current pieces in order to be able to make a proposal.¹⁰

⁸See example 36.

⁹I also suggested an additional function of IRUs as a *WARRANT*; it is not necessary to reason about whether to accept a *WARRANT* that is already mutually believed.

¹⁰To realize the goals of the simulation, it would not make sense to put limitations on agents' utterance producing capability since this would require a theory of suboptimal utterance production and could potentially involve the design of a number of strategies for recovery from suboptimal utterances. Limits on the amount of time spent deliberating will automatically follow from limits on means-end reasoning since fewer options will be passed to the deliberation component.

Section 8.5.3 discussed the fact that there is an interaction between Consequence and Attention. One case of this interaction will affect the simulation with respect to inferences about the achievement of MATCHED-PAIR goals. In the Design-World task the subgoal of a MATCHED-PAIR can be simultaneously satisfied by an action that contributes to a (Design ?Room) goal. Thus it is possible that an agent would only infer the (Design ?Room) goal and not the MATCHED-PAIR goal. However, an explicit segment opening statement can function to obviate the need to make an inference that the current proposal can contribute to a MATCHED-PAIR subgoal.

8.5.2.1 Prediction

An agent embodying the explicit Consequence strategy will achieve better scores than an agent embodying the implicit Consequence strategy when interacting with an agent that is limited as time allowed for the performance of means-end reasoning.

8.5.3 Attention Strategies

There will be two Attention strategies. However, since any IRU, and indeed any utterance, manipulates attentional state when it is said, there is an overlap between Attention and the other two communicative functions. Because of this, the Attention strategies will be as independent as possible of Attitude and Consequence. The explicit Attention strategy IRUs will consist of statements to open and close discourse segments.

The concept of Attention in this account draws on Grosz and Sidner's notion of focus spaces (Grosz, 1977; Sidner, 1979; Grosz and Sidner, 1985). According to their theory, each discourse segment has a corresponding focus space containing the relations, propositions and discourse entities evoked in that segment. Since this is a fairly simple task, discourse segments will arise naturally from the task structure, with a segment corresponding to each subgoal of the task and the structure of the task driving the structure of the discourse (Hobbs and Evans, 1979; Power, 1974). Each segment consists of a number of propositions including (1) the current goal; (2) the current proposal; and (3) other related propositions evoked during the dialogue corresponding to the current segment. Propositions will be linked by one of the propositional relations discussed earlier in 5.1. The examples given below will provide motivation for a discussion as to whether the concept of a focus space requires a stack mechanism.

The hypothesis is that the IRU in 47-(17) functions to close the segment starting with the question given in 47-(6).

- (47) (6) r. or uh *does that income from the certificate of deposit rule her out as a dependent*
 (7) h. *yes it does*
 (8) r. *it does*
 (9) h. *yup, that knocks her out.*
 now there is something you can do. do you support her in any way?
 (10) r. *yes i mean she yeah we supply everything, heat, light, food, in other words we*

you know she pays nothing as far as the uh upkeep of the home
 (11) h. the only amount you have spent then is indirect.
 r. uh
 h. there's nothing direct
 r. yea
 h. then that you spend
 (12) r. food
 (13) h. the rest of her support comes out of her three thousand and social security
 (14) r. yeah whatever clothes she needs er or uh dental care, that kind of thing
 (15) h. well the medical and dental care you can deduct provided you can establish
 that you have provided more than half support.
 (16) r. uh huh
 (17) h. BUT THE DEPENDENCY YOU CANNOT CLAIM
 (18) r. um hm (breath) I see.
 ok. uhh, alright, the second question...

The fact that 47-(17) occurs is an argument for the existence of some kind of stack mechanism associated with discourse segments. The intervening discussion from (9) to (16) seems to be carried out with an understanding that the question given in (6) is still pending in some sense. In the focus space model this would mean that (9) to (16) is an embedded segment. Utterance 47-(17) indicates that that embedded segment is closed as well as its parent segment corresponding to the dialogue up to (9).

Discourses can be opened with IRUs that set the context for what the discourse will be about. For example, if one American said to another on November 10th that *Thanksgiving is coming up*, that utterance could hardly be said to be informative by itself. It would probably serve to set the context however, and to simplify the process of making sense of a following utterance such as *Have you bought a turkey?*. In the HG corpus, I didn't count beliefs that were common cultural knowledge as mutual beliefs, in order to ensure that the IRUs that this analysis is based on really consist of information that should have been mutually believed, and in order to make my criteria for selecting IRUs replicable. The only examples of segment opening statements that could possibly be counted as IRUs under my criteria would be those whose propositional material had already been discussed in the same dialogue. Therefore the examples given below consist of 'reopening' a discourse segment.

These discourse segment reopenings often have the sense of 'picking up where you left off'. A segment may have been left suspended for a number of reasons such as an interruption by another speaker or a need to temporarily turn to another topic. Consider the utterance in 48-(23).

- (48) (13) e. Well however this is my question: I am a single woman, and retired. I have about 120 M, and I'll tell you how it's broken down and perhaps you can advise me
 (14) h. How old are you elsa?
 (15) e. I am seven six, alright?
 (16) h. I got you
 (17) e. uh 70,000 in CDs, 30,000 -

- (18) h. When are they due?
 (19) e. Pardon?
 (20) h. When are they due?
 (21) e. Well they're due right now, every month one is due until June
 (22) h. They're due one a month?
 (23) e. Yes sir. NOW THAT'S 70.
 Now another 30 in low-income CDs at 8% – the long term ones, you know?
 (24) h. When are they due?

After an interruption by Harry, starting at (18), Elsa (e), in (23), tries to continue her enumeration of how her money is invested. When she starts this, she first produces a summary of the extent of her investments she had accounted for so far at the time of the interruption. The IRU marks the re-opening of the previous segment. It would have been possible for this IRU to simultaneously involve Consequence as well as Attention. For instance if the interruption from Harry had come after Elsa (e) had described how she had an additional 30 thousand of her money invested, Elsa might have said *Now that's 100* instead of *Now that's 70*. In either case it would have marked a return to the previous segment. But in the former case it would have also made an inference explicit.

Another way in which Attention interacts with Consequence is that an IRU that closes a segment might make an inference explicit, rather than just repeating or paraphrasing information. Consider utterance (26) in the excerpt from example 45 repeated here for convenience:

- (49) (20) h. right. The maximum amount of credit that you will be able to get will be 400 that *THEY will be able to get will be 400 dollars on their tax return*
 (21) c. *Four hundred dollars for the whole year?*
 (22) h. *Yeah it'll be 20%*
 (23) c. um hm
 (24) h. Now if indeed they pay the two thousand dollars to your wife, that's great.
 (25) c. um hm
 (26) h. SO WE HAVE FOUR HUNDRED DOLLARS. Now as far as you are concerned that could cost you more. Remember, you're gonna have 2000 worth of income. What's your tax bracket?

It would be possible to have an agent that was both attention-explicit and also consequence-explicit produce IRUs like this rather than the simple closing messages discussed below. The simulation will explore this interaction between Attention and Consequence.

The basic component of the explicit Attention strategy consists of statements that indicate the opening and closing of a discourse segment. The implicit Attention strategy, as a control, will not include these opening and closing statements. Opening and closing statements for the simulation are given below:

(Say A B
 (Open (Achieve A&B (Design Room-1))))

(Say A B
 (Close (Achieve A&B (Design Room-1))))

These messages are a poor match for the diversity of surface forms that opening and closing statements have in naturally occurring dialogues, but the interpretation of language forms is not the topic of this dissertation and so much of this diversity will have to remain a topic for future work. However, once the first round of simulations are done, testing some of the basic predictions, it may be possible to test Attention strategies that directly interact with Consequence. Example in Design-World would be a closing statement such as *That gets us 120 points so far* or a (re)-opening statement such as *We agreed on putting the red-chair and red-couch in room-1*, when returning to a goal such as the (Design Room-1).

Opening and closing statements are equivalent to Carletta's low-risk context articulation strategy (Carletta, 1992). Carletta states that this strategy in actuality has no effect on agents' knowledge because the statements about the task itself always make the goal explicit. This is generally the case here as well. For example a proposal to (Put A&B Red-Couch Room-1 Later) can be easily inferred to be an action that CONTRIBUTES to the goal (Achieve A&B (Design Room-1)). There is no need to have this goal made explicit beforehand. Thus these opening statements are essentially IRUs that make these inferences explicit.¹¹ Even though agents can make these inferences, it may be that in this simulation environment, unlike Carletta's, the difference between Attention Strategies will interact with resource-bounds and the calculation of costs associated with completing the task.

Furthermore, there seems to be more going on in the naturally occurring dialogues than just indicating the opening of a segment or marking the reopening of a previous segment. An excellent example is provided by the following dialogue excerpt, repeated here from chapter 6:

- (50) (3) e.
 – and i was wondering – should i continue on with the certificates or
 (4) h. well it's difficult to tell because we're so far away from any of them – but i would suggest this – if *all of these are 6 month certificates and i presume they are*
 (5) e. *yes*
 (6) h. *then i would like to see you start spreading some of that money around*
 (7) e. uh hu
 (8) h. now in addition, how old are you
 .
 (discussion and advice about starting an IRA)
 .
 (21) e. uh huh and
 (22) h. but as far as the certificates are concerned, I'D LIKE THEM SPREAD OUT A LITTLE BIT - THEY'RE ALL 6 MONTH CERTIFICATES
 (23) e. yes
 (24) h. and i don't like putting all my eggs in one basket - and i would suspect that

¹¹In the distribution analysis of the HG corpus, I wouldn't count utterances that state the goal as IRUs since it wouldn't be clear that they would have been in that situation.

february 25 would be a good time to put it into something that runs for 2 and a half years. that first one that comes due. call me on the others as they come due

I would particularly like to focus on 50-(22). I repeat (22) here labelling the clauses under discussion:

(51)22a.but as far as the certificates are concerned,

22b. I'D LIKE THEM SPREAD OUT A LITTLE BIT -

22c. THEY'RE ALL 6 MONTH CERTIFICATES

It cannot be the case that the only purpose of the IRUs in 51-22b and 51-22c is to evoke a previous context; 51-22a alone should be adequate to mark a return to a previous segment in which *certificates* were being discussed. Yet, if the focus space representation of that previous segment includes all the propositions and discourse entities that were part of it (Grosz, 1977; Sidner, 1979), how could these IRUs possibly have a function? If we assume that the speaker here is rational, it would seem that they must have an additional function related to a return to a previous segment. One explanation would be that not all of the previously discussed items are currently relevant, and the function of the IRU is to select as the locus of attention some limited subset of the items in the previous segment. Another explanation would be that by default none of the items in the previous segment are salient and they must be re-evoked. However, Grosz's naturally occurring examples of a felicitous reference to *the pump* as *it*, after the pump hadn't been mentioned for a long period of time would seem to contradict this. It is possible that another explanation for that could be found. For instance, it might be worth considering the possibility that the pump in that situation was SITUATIONALLY EVOKED (Prince, 1981b).

Under either of these hypotheses, the open question is exactly how a speaker (and in this case a Design-World artificial agent), decides which items in the focus space of a previous segment should be restated in some form in the current segment. There are two separate issues here: (1) whether or not **any** previous material is restated, and (2) what previous material is restated. I will refer to (1) as the REINVOCATION problem and (2) as the SELECTION problem. These are distinct problems. The Attention strategy distinguishes between agents who do or do not reinvoke previous material. This leaves aside the selection problem. Different algorithms could be involved in selection. For instance, **all** the items in a previous segment could be selected, or items could be selected with a simple strategy such as most recently used. Another selection strategy would be select only certain types of items from the previous segment that are the basis of lots of inferences such as Points propositions. I will start with a simple selection strategy of randomly selecting a proposition to repeat, and hope to refine this as the simulation progresses.

Although there is evidence that attention/working memory capacity is limited in human agents (Miller, 1956), I will not actually place a limit on agents' attentional capacity. I expect that these limitations will show up in the time to retrieve a belief that will follow from the attention/memory model incorporated into the simulation. This model must incorporate both frequency and recency effects for beliefs discussed during a dialogue (Landauer, 1975). The time or number of steps required to retrieve beliefs will be kept track of as a cost in performing the task.

8.5.3.1 Prediction

The prediction is that an agent embodying an explicit Attention strategy will achieve higher task scores than an agent embodying an implicit Attention strategy, when the cost of retrieving prior beliefs and intentions is deducted from the score achieved.

Another prediction is that the explicit Attention strategy may interact with agents that are limited means-end reasoners by reducing the amount of inference necessary to infer a matched-pair goal.

8.5.4 Summary

This section has discussed the communicative strategies that will be used in the simulation. Initial tests of the hypotheses will be carried out using two strategies for each of Attitude, Consequence and Attention. These strategies will allow comparisons between explicit and implicit realization of each of the putative IRU communicative functions.

The strategies will interact with communication channel and agent properties of noise, limited inference and attention. While the potential space for the simulation is quite large, I will only explore certain points in the space where I would expect particular interactions to occur.

The model of limited attention needs to be refined since it will be a major determinant of the way the simulation works. Limited attention may interact with limited inference and in fact could be the cause of agents' limited inferential capabilities.

8.6 Comparison with Other Work

While there is certainly a great deal of previous research in linguistics, philosophy, psychology and computational linguistics that is related to this work, this review focuses on computational models of dialogue or multi-agent communication that are especially relevant to the approach taken here.

The approach adopted in this thesis originates with the work of Power (Power, 1974; Power, 1984). Power was the first person to test a theory of dialogue through an experimental system with two artificial agents who collaborated to achieve their goals. Power's system was mixed initiative, and control was alternated between agents. My work differs from Power's in focusing on a particular class of dialogue phenomena and in trying to demonstrate the utility of particular dialogue strategies as compared to others.

The most recent work in the same vein as Power is that of Carletta (Carletta, 1992). Carletta explores different styles of communication in the context of a Map-Task Dialogue, where the goal of the task is for the planning agent, the instructor, to instruct the reactive agent, the instructee, how to get from one place to another on the map. The instructor has a map with a number of landmarks and a route on it. The instructee's map's landmarks may not exactly match. The instructee must draw the route on their version of the map. Carletta's simulated dialogues are single initiative, which follows directly from the type of task that is

being modeled. The instructor embodies a number of different kinds of strategies for how to accomplish the goal of getting the instructee to draw a route on a map.

Carletta’s work explores many issues that are relevant for this thesis, including ‘high risk’ dialogue strategies, and methods of recovering from these strategies when failure occurs. Agents can take risks with respect to whether or not they make the context explicit, what Carletta calls ‘context articulation’. This is similar to the explicit Attention strategy discussed above. However Carletta notes that, in the Map task, context articulation doesn’t serve any function because the context is always obvious. A more critical ‘high risk’ strategy involves definite references to locations that may not appear on the instructee’s map, e.g. *the big river*. The ‘low risk’ version of achieving the same goal might involve asking the instructee whether or not they have *a big river* on their map, before producing an instruction that refers to it. High risk strategies produce failures in communication that must be recovered from. The only actions that the instructee can initiate are clarifications to address referential failures produced by high risk strategies. The instructor then initiates one of several recovery strategies, using a notion of utility to select between them. The recovery strategy selected is based on calculations of the expected cost of repairing the current plan or replanning from scratch.

Carletta argues that ‘high risk’ strategies are more efficient, but there are problems with evaluating the various strategies in the Map Task. This is because it is difficult to define a performance measure for the Map Task, and the subjects who carried out the task were not told that their performance would be measured by any particular criteria. The only measure of efficiency possible is the length of the dialogue, but this ignores the quality of the solution, and indeed some of the subject pairs produced routes that barely approximated the one they were supposed to draw. This means that the Map Task dialogue corpus cannot be used to explore the efficiency trade-offs of different dialogue strategies.

My work differs from Carletta’s in focusing on mixed-initiative dialogues, and using a task in which both agents propose and accept plans. In addition, the fact that the value of an accepted plan is easily calculable means that I have an objective measure of the success or failure of various strategies. In both the human-human dialogues that I intend to collect and in the artificial agent dialogues, there is a clear notion of maximizing utility of the collaborative plan.

The notion of a collaborative plan is very close to that of Grosz and Sidner’s (G&S) Shared-Plan (Grosz and Sidner, 1990). However, because there are some contrasts, I have proposed my own definitions for the sake of simplicity. I use a simpler plan representation than the one that they have proposed because my focus is on the dialogue behavior that produces the plan, rather than the plan itself. I assume that the execution of all actions is intended by both agents. This simplification is possible because the proposed actions are not actually executed by robots. If they were to be executed, then which agent is to do what when would also have to be discussed. Presumably, if one agent were capable of putting a piece of furniture in a room, there would be no need for two agents to synchronize on this activity. I also incorporate a utility measure as part of the definition of a collaborative plan, and assume a global utility function defined by the Design-World task.

Another difference between the two accounts is in the treatment of coming to an agreement to carry out a proposed action. According to G&S’s Conversational Default Rule (CDR2), the non-initiating agent will adopt the initiating agent’s goals as a default:


```

SharedPlan* (G1,G2 Achieve P)) & Desire (G1, Do (G2, action))
& Believe (G2, Exec (G2, Action)) & Believe (G2, Contribute (Action, (Achieve
P)))
→ Intend (G2, Action)

```

This default rule licenses the inference of mutual beliefs based on the previous mental state of the hearer, but not on any evidence the hearer provides(cf. (Perrault, 1990)). The only exception to this is in the first utterance of the dialogue where the non-initiating agent agrees to construct a SharedPlan. Once the non-initiator has agreed to construct a SharedPlan, it seems that she has also agreed to follow the strategy that the initiator of the SharedPlan proposes. In contrast, in the dialogue situation that I explore here, the conversants start out with a mutual goal to achieve a design for the floor plan, but cannot be sure that the other conversant will either understand or agree with everything they propose.

Another difference is that because G&S are modeling collaboration in the domain, they have no need to model understanding. One consequence of this is that utterances such as *O.K.* add beliefs about executability directly. In my account, an utterance such as this contributes to the endorsement on the attention assumption that underlies the mutual belief about understanding. The belief about executability follows as an inference from the collaborative principle. Recent work in progress by Sidner develops the SharedPlan account with a *negotiation* language(Sidner, 1992). Sidner's recent work hasn't yet addressed issues of resource-bounds and efficient strategies. In addition, this approach doesn't make clear under what circumstances agents make, accept and reject proposals, or when they ask for or provide additional information to support decisions.

This work is also related to research by Galliers (Galliers, 1989; Galliers, 1991b; Galliers, 1991a). In Galliers' theory of communication, conflict is a beneficial and necessary component of cooperation. Galliers theory includes a model of belief revision, ABR, that supports agents' decisions about whether and how they wish to revise their beliefs in the light of new incoming information. ABR is the account of belief that I assume here and is the basis for the account of mutual beliefs that I have developed as well. Recently this group has been carrying out a number of empirical evaluations of their framework. Their focus has been on architectural variations that affect how many inter-agent conflicts are resolved in the course of the dialogue(Cawsey et al., 1992). My focus on resource-bounds and associated communicative strategies distinguishes my work from theirs.

This work also is related to that of Litman(Litman and Allen, 1990), who relates discourse intentions to task plans. My treatment of discourse plans is not as sophisticated as Litman's. I simplify the simulation here by using schemas in which achieving mutual belief is part of the domain defined task. The need to achieve mutual belief is what motivates communication and agents in this simulation communicate in fairly stereotypical patterns. Another difference between this work and theirs is captured by Litman and Allen's INFORM AXIOM:

```

Header: Inform (S, H, P)
Prerequ: Know (S, P)
Decomp: Surface-Inform (S,H,P)

```

Effects: Know (H, P), KNOW (H, KNOW (S, P))

This stipulates that after an inform action the hearer knows the proposition that the speaker intended to convey. My work extends this approach by considering cases where this process can go awry, for instance cases where the hearer does not understand what the speaker intended to convey, or cases where the hearer does not wish to adopt a proposal put forth by the speaker.

Another point of divergence with the accounts of Litman, Allen, Cohen and Perrault is in the existence of IRUs themselves. Their speech-act based accounts of plan recognition and generation would have to be extended in some way to account for the presence of IRUs. Perrault has proposed that the recognition of redundancy actually triggers the inference of an indirect speech act, but many of the IRUs presented here can be understood as INFORMS despite the fact that the information they convey is already in the representation of mutual beliefs. Cohen's speech act generation axioms rule out the generation of INFORM utterances whose propositional content has already been conveyed, however IRUs could be produced if a new speech act was introduced. Litman and Allen's plan inference heuristics disprefers the recognition of a plan whose effect has already been achieved. This means that IRUs would not be recognized as INFORM speech acts, and again either the definition of an INFORM would have to be changed or a new speech act would have to be introduced.

Figure 8.6 summarizes the discussion above on the relationship of this work to other computational models with respect to certain agent/dialogue properties. The contributions of this thesis are in exploring the relationship between agents' cognitive properties and certain types of dialogue behavior. I focus on mixed-initiative problem-solving dialogues, between resource-bounded autonomous agents.

Agent Property	GS90	P74	C92	LA90	G89	This Thesis
Autonomous	?	N	N	N	Y	Y
Resource-Bounded	N	N	N	N	N	Y
Own Goals	?	Y	N	N	Y	Y
Own Preferences	N	N	N	N	Y	Y
Own Bels	Y	Y	Y	N	Y	Y
Achieve MBs	Y	?	Y	Y	N	Y
Misunderstanding	N	N	Y	Y	N	Y
Produce IRUs	N	N	N	N	N	Y

Figure 8.8: Agent Properties in Other Work

KEY

GS90: (Grosz and Sidner, 1990)

P74: (Power, 1974)

C92: (Carletta, 1992)

LA90: (Litman and Allen, 1990)

G89: (Galliers, 1989; Galliers, 1991b; Galliers, 1991a)

The dialogue model presented here assumes that agents are autonomous. Previous work has often assumed that agents are cooperative and helpful and will adopt the goals and beliefs of others (but see (Galliers, 1989)). I have dropped this simplifying assumption in order to motivate utterances in which agents explicitly indicate acceptance. I have introduced parameters into the simulation model in order to be specific about when agents will adopt the goals and beliefs of others and when they won't, in order to explore the trade-offs of different strategies when agents are autonomous.

Both of the agents engaged in the dialogue model presented here have their own goals, preferences and beliefs. Dialogues are simpler when one agent's beliefs are a subset of the other's (Rosenschein, 1985), and sometimes previous researchers have made this simplification.

The agents in the simulation produce IRUs in particular situations. A dialogue strategy that includes IRUs would appear to be a patently inefficient strategy. One goal of the simulation is to see whether IRUs can actually reduce processing in different discourse situations. This focus on resource-bounds distinguishes this work from previous accounts of dialogue.

Agents' Attitude strategies are hypothesized to relate to the capability for misunderstanding. Some previous research has focused on how agents recover from misunderstanding, but no-one to my knowledge has attempted to demonstrate the computational efficiency of strategies which avoid misunderstanding in the first place.

8.7 Summary

This proposal has presented a class of INFORMATIONALLY REDUNDANT utterances that violate in one respect or another previous accounts of dialogue. I have classified these IRUs according to three communicative functions: Attitude, Consequence and Attention. I have argued that these utterances reflect the fact that human agents are resource-bounded in particular ways.

I haven't argued that IRUs are the only way that these functions can be achieved, only that they seem to be a natural way to achieve these functions in the circumstances in which they occur. In addition, it is sometimes difficult to tell whether the primary speaker intention in producing an utterance is related to only one of the above functions. Individual utterances often seem to function at more than one level, since for instance, every utterance manipulates the attention of the discourse participants.

Of course other accounts of the phenomena discussed here are possible, and there are sure to be ways of dealing with the utterances discussed here within other theories of dialogue.

I have conducted the initial phases of a distributional analysis over a large corpus of naturally occurring advisory dialogues in an attempt to validate some of my claims. I also argue that my claims about processing are difficult to validate with a corpus analysis and have proposed a computational experiment that should provide further support for my claims. This experiment should also help to tease apart the relationship between an agent's intentions and the surface form of the utterances that they produce.

8.8 Acknowledgments

I'd like to thank my advisors, my committee members and my colleagues for many useful discussions and helpful comments: Aravind Joshi and Ellen Prince have advised me in this work. My committee of Karen Sparck Jones, Mark Liberman and Bonnie Webber have all provided many useful comments on the ideas here. Lastly thanks to my colleagues and friends: Susan Brennan, Janet Cahn, Jean Carletta, Alison Cawsey, Herb Clark, Sharon Cote, Barbara Di Eugenio, Dave Frohlich, Nick Haddock, Julia Galliers, Ellen Germain, Masayo Iida, Libby Levison, Cynthia McLemore, Megan Moser, Christine Nakatani, Owen Rambow, Steve Reece, Phil Resnik, Craige Roberts, Penni Sibun, Candy Sidner, Phil Stenton, and Steve Whittaker.

Bibliography

- James F. Allen. 1979. A plan-based approach to speech act recognition. Technical report, University of Toronto.
- J. R. Anderson and G. H. Bower. 1973. *Human Associative Memory*. V.H. Winston and Sons.
- J. R. Anderson. 1974. Verbatim and propositional representation of sentences in immediate and long term memory. *Journal of Verbal Learning and Verbal Behavior*, 13:149–162.
- Cecile Balkanski. 1990. Modelling act-type relations in collaborative activity. Technical Report TR-23-90, Harvard University.
- Jon Barwise. 1988. The situation in logic-iv: On the model theory of common knowledge. Technical Report No. 122, CSLI. Chapter 8 of *The situation in Logic*, also as *Three Views of Common Knowledge in TARK2*.
- Sara S. Bly. 1988. A use of drawing surfaces in different collaborative settings. In *Proceedings of the Conference on CSCW*.
- J. D. Bransford, J. R. Barclay, and J. J. Franks. 1972. Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 3:193–209.
- Michael Bratman, David Israel, and Martha Pollack. 1988. Plans and resource bounded practical reasoning. *Computational Intelligence*, 4:349–355.
- Susan E. Brennan. 1990. *Seeking and Providing Evidence for Mutual Understanding*. Ph.D. thesis, Stanford University Psychology Dept. Unpublished Manuscript.
- Derek G. Bridge. 1991. *Computing Presuppositions in an incremental natural language processing system*. Ph.D. thesis, Cambridge University. Technical report number 237, Cambridge Computer Lab.
- Jean C. Carletta. 1992. *Risk Taking and Recovery in Task-Oriented Dialogue*. Ph.D. thesis, Edinburgh University.
- Alison Cawsey, Julia Galliers, Steven Reece, and Karen Sparck Jones. 1992. Automating the librarian: A fundamental approach using belief revision. Technical Report 243, Cambridge Computer Laboratory.
- Herbert H. Clark and Susan E. Brennan. 1990. Grounding in communication. In L. B. Resnick, J. Levine, and S. D. Bahrend, editors, *Perspectives on socially shared cognition*. APA.
- Herbert H. Clark and Eve V. Clark. 1977. *Psychology and Language*. Harcourt Brace Jovanovich, Inc.
- Herbert H. Clark and Susan E. Haviland. 1977. Comprehension and the given-new contract. In R. O. Freedle, editor, *Discourse Production and Comprehension*, pages 1–40. Ablex Publishing, Norwood, N.J.

- Herbert H. Clark and Catherine R. Marshall. 1981. Definite reference and mutual knowledge. In Joshi, Webber, and Sag, editors, *Elements of Discourse Understanding*, pages 10–63. CUP, Cambridge.
- Herbert H. Clark and Edward F. Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13:259–294.
- Herbert H. Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22:1–39.
- Phillip R. Cohen and Hector Levesque. 1990. Rational interaction as the basis for communication. In *Cohen, Morgan and Pollack, eds. Intentions in Communication*, MIT Press.
- Phil Cohen and Hector Levesque. 1991. Teamwork. *Nous*, 25:11–24.
- Phillip R. Cohen, Hector J. Levesque, Jose H. T. Nunes, and Sharon L. Oviatt. 1990. Task oriented dialogue as a consequence of joint activity. In *Pacific Rim Conference on Artificial Intelligence*.
- Phillip R. Cohen. 1978. On knowing what to say: Planning speech acts. Technical Report 118, University of Toronto; Department of Computer Science.
- Robin Cohen. 1987. Analyzing the structure of argumentative discourse. *Computational Linguistics*, 13:11–24.
- Cruttenden. 1986. *Intonation*. Cambridge University Press.
- J. de Kleer. 1986. An assumption-based tms. *Artificial Intelligence*, 28:127–162.
- Barbara Di Eugenio. 1992. Goals and Actions in Natural Language Instructions. Technical Report MS-CIS-92-07, University of Pennsylvania. Dissertation Proposal.
- Julia R. Galliers. 1989. A theoretical framework for computer models of cooperative dialogue, acknowledging multi agent conflict. Technical Report 172, University of Cambridge, Computer Laboratory.
- Julia R. Galliers. 1990. Belief revision and a theory of communication. Technical Report 193, University of Cambridge, Computer Laboratory, New Museums Site, Pembroke St. Cambridge England CB2 3QG.
- Julia R. Galliers. 1991a. Autonomous belief revision and communication. In P. Gardenfors, editor, *Belief Revision*, page 1. Cambridge University Press.
- Julia R. Galliers. 1991b. Cooperative interaction as strategic belief revision. In M.S. Deen, editor, *Cooperating Knowledge Based Systems*, page 1. Springer Verlag.
- S. Garrod and T. Trabasso. 1973. A dual-memory information processing interpretation of sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, 12:155–167.
- Gerald Gazdar. 1979. *Pragmatics : implicature, presupposition and logical form*. Academic Press.

- H. P. Grice. 1957. Meaning. *Philosophical Review*, LXVI, No. 3:377–388.
- H. P. Grice. 1967. *William James Lectures*.
- H. P. Grice. 1975. Logic and conversation. In D. Davidson and G. Harman, editors, *The Logic of Grammar*. Dickenson Publishing Co.
- Barbara J. Grosz and Candace L. Sidner. 1985. The structure of discourse structure. Technical Report CSLI-85-39, Center for the Study of Language and Information, Stanford, CA.
- Barbara J. Grosz and Candace L. Sidner. 1986. Attentions, intentions and the structure of discourse. *Computational Linguistics*, 12:175–204.
- Barbara J. Grosz and Candace L. Sidner. 1990. Plans for discourse. In *Cohen, Morgan and Pollack, eds. Intentions in Communication*, MIT Press.
- Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1983. Providing a unified account of definite noun phrases in discourse. In *Proc. 21st Annual Meeting of the ACL, Association of Computational Linguistics*, pages 44–50.
- Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1986. Towards a computational theory of discourse interpretation. Unpublished Manuscript.
- Barbara J. Grosz. 1977. The representation and use of focus in dialogue understanding. Technical Report 151, SRI International, 333 Ravenswood Ave, Menlo Park, Ca. 94025.
- M.A.K. Halliday and Ruquaiya Hasan. 1976. *Cohesion in English*. Longman.
- Gilbert Harman. 1986. *Change of View*. MIT PRESS.
- Julia Hirschberg. 1985. *A Theory of Scalar Implicature*. Ph.D. thesis, University of Pennsylvania, Computer and Information Science.
- J.R. Hobbs and D. Evans. 1979. Conversation as planned behaviour. Technical Report 203, Artificial Intelligence Center, SRI International, Menlo Park, CA.
- Jerry R. Hobbs and Paul Martin. 1987. Local pragmatics. Technical report, SRI International, 333 Ravenswood Ave., Menlo Park, Ca 94025.
- Jerry R. Hobbs. 1979. Coherence and coreference. *Cognitive Science*, 3:67–90.
- Jerry R. Hobbs. 1985. On the coherence and structure of discourse. Technical Report CSLI-85-37, Center for the Study of Language and Information, Ventura Hall, Stanford University, Stanford, CA 94305.
- Beth A Hockey. 1991. The function of okay as a cue phrase in discourse. In *AAAI Fall symposium on Discourse Structure*, November. Department of Linguistics, University of Pennsylvania.
- Laurence R. Horn. 1984. Toward a new taxonomy for pragmatic inference: q-based and r-based implicature. In D. Schiffrin, editor, *Meaning Form and Use in Context: Linguistic Applications*, pages 11–42. Georgetown University Press.

- Laurence R. Horn. 1991. Given as new: When redundant affirmation isn't. *Journal of Pragmatics*, 15:305–328.
- Aravind K. Joshi, Bonnie Lynn Webber, and Ralph M. Weischedel. 1984. Preventing false inferences. In *COLING84: Proc. 10th International Conference on Computational Linguistics.*, pages 134–138.
- Aravind K. Joshi, Bonnie Lynn Webber, and Ralph M. Weischedel. 1986. Some aspects of default reasoning in interactive discourse. Technical Report MS-CIS-86-27, University of Pennsylvania, Department of Computer and Information Science.
- Aravind K. Joshi. 1982. Mutual beliefs in question-answer systems. In Neil V. Smith, editor, *Mutual Knowledge*, pages 181–199. Academic Press, New York, New York.
- Marcel A. Just and Patricia A. Carpenter. 1992. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1).
- Daniel Kahneman, Paul Slovic, and Amos Tversky. 1982. *Judgment under uncertainty : heuristics and biases*. Cambridge University Press.
- Henry Kautz. 1990. A circumscriptive theory of plan recognition. In *Cohen, Morgan and Pollack, eds. Intentions in Communication*, MIT Press.
- Alison Kidd. 1985. The consultative role of an expert system. In P. Johnson and S. Cook, editors, *People and Computers: Designing the Interface*. Cambridge University Press, Cambridge, U.K.
- Kurt Konolige. 1985. Belief and incompleteness. In Jerry R. Hobbs and Robert C. Moore, editors, *Formal Theories of the Commonsense World*, pages 359–403. Ablex Publishing.
- Robert Ladd. 1980. *The Structure of Intonational Meaning: Evidence from English*. University of Indiana Press. A Cornell University Dissertation.
- Thomas K. Landauer. 1975. Memory without organization: Properties of a model with random storage and undirected retrieval. *Cognitive Psychology*, pages 495–531.
- Hector J. Levesque, Phillip R. Cohen, and Jose H. T. Nunes. 1990. On acting together. In *AAAI90*.
- Stephen C. Levinson. 1979. Activity types and language. *Linguistics*, 17:365–399.
- Stephen C. Levinson. 1981. Some pre-observations on the modelling of dialogue. *Discourse Processes*, 4:93–116.
- Stephen C. Levinson. 1983. *Pragmatics*. Cambridge University Press.
- David Lewis. 1969. *Convention*. Harvard University Press.
- Mark Y. Liberman and Cynthia A. McLemore. 1992. The structure and intonation of business telephone openings. *The Penn Review of Linguistics*, 16:68–83.

- Mark Y. Liberman and Ivan A. Sag. 1974. Prosodic form and discourse function. In *CLS 10*, pages 416–27.
- Mark Y. Liberman. 1975. *The Intonational System of English*. Ph.D. thesis, MIT. (published by Garland Press, NY, 1979).
- Diane Litman and James Allen. 1990. Recognizing and relating discourse intentions and task-oriented plans. In *Cohen, Morgan and Pollack, eds. Intentions in Communication, MIT Press*.
- Kathleen R. McKeown. 1985. Discourse strategies for generating natural language text. *Artificial Intelligence*, 27(1):1–42, September.
- Cynthia A. McLemore. 1991. *The pragmatic Interpretation of English Intonation: Sorority Speech*. Ph.D. thesis, University of Texas, Austin.
- Igor A. Melčuk. 1988. *Dependency Syntax: Theory and Practice*. SUNY.
- G. A. Miller. 1956. The perception of speech. In M. Halle, editor, *For Roman Jakobson*, pages 353–359. The Hague, Mouton Publishers.
- Gopalan Nadathur and Aravind K. Joshi. 1983. Mutual beliefs in conversational systems: Their role in referring expressions. In *Proc. International Joint Conference on Artificial Intelligence, Austin*.
- Josef Perner and Alan Garnham. 1988. Conditions for mutuality. *Journal of Semantics*, 6:369–385.
- Ray Perrault. 1990. An application of default logic to speech-act theory. In *Cohen, Morgan and Pollack, eds. Intentions in Communication, MIT Press*.
- Martha E. Pollack and Marc Ringuette. 1990. Introducing the tileworld: Experimentally evaluating agent architectures. In *AAAI90*, pages 183–189.
- Martha Pollack, Julia Hirschberg, and Bonnie Webber. 1982. User participation in the reasoning process of expert systems. In *AAAI82*.
- Martha Pollack. 1986. Inferring domain plans in question answering. Technical Report 403, SRI International - Artificial Intelligence Center. University of Pennsylvania Dissertation.
- Martha E. Pollack. 1990. Plans as complex mental attitudes. In *Cohen, Morgan and Pollack, eds. Intentions in Communication, MIT Press*.
- Martha E. Pollack. 1991. The uses of plans. In *Talk given on receipt of the Computers and Thought Award at AAAI91*.
- Richard Power. 1974. *A Computer Model of Conversation*. Ph.D. thesis, University of Edinburgh.
- Richard Power. 1979. *Linguistics*.

- Richard Power. 1984. Mutual intention. *Journal for the Theory of Social Behaviour*, 14.
- Ellen F. Prince. 1978. On the function of existential presupposition in discourse. In *Papers from 14th Regional Meeting*. CLS, Chicago, IL.
- Ellen F. Prince. 1981a. Topicalization, focus movement and yiddish movement: a pragmatic differentiation. In D. Alford et al., editor, *Proceedings of the Seventh Annual Meeting of the Berkeley Linguistics Society*, pages 249–264. BLS.
- Ellen F. Prince. 1981b. Toward a taxonomy of given-new information. In *Radical Pragmatics*. Academic Press.
- Ellen F. Prince. 1985. Fancy syntax and shared knowledge. *Journal of Pragmatics*, pages 65–81.
- Ellen F. Prince. 1986. On the syntactic marking of the presupposed open proposition. *CLS86*.
- Michael Reddy. 1979. The conduit metaphor – a case of frame conflict in our language about language. In A. Ortony, editor, *Metaphor and Thought*, pages 284–324. Cambridge University Press, Cambridge.
- Jeffrey S. Rosenschein. 1985. *Rational Interaction and Multi Agent Planning*. Ph.D. thesis, Stanford University Computer Science Dept.
- Robert Rubinoff. 1987. Scalar implicature and the given/new distinction. In *Penn Linguistics Colloquium*.
- Harvey Sacks, Emmanuel Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking in conversation. *Language*, 50:325–345.
- Jerrold M. Sadock. 1978. On testing for conversational implicature. In Peter Cole, editor, *Syntax and Semantics, Volume 9: Pragmatics*, pages 281–297. Academic Press.
- Emanuel A. Schegloff. 1982. Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In D. Tannen, editor, *Analyzing Discourse: Text and Talk*, pages 71–93. Georgetown University Press.
- Emanuel A. Schegloff. 1987. Between micro and macro: Contexts and other connections. In *The Micro-Macro Link*. University of California Press.
- Stephen R. Schiffer. 1972. *Meaning*. Clarendon Press.
- Deborah Schiffrin. 1982. Cohesion in everyday discourse: The role of paraphrase. *Sociolinguistics Working Paper*, 97.
- Deborah Schiffrin. 1987. *Discourse Markers*. Cambridge University Press.
- Marcel Schoppers. 1988. *Representation and Automatic Synthesis of Reaction Plans*. Ph.D. thesis, University of Illinois at Urbana-Champaign. Draft.
- John R. Searle. 1975. Indirect speech acts. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics III: Speech Acts*, pages 59–82. Academic Press, New York.

- Candace Sidner and David Israel. 1981. Recognizing intended meaning and speakers plans. In *Proc. International Joint Conference on Artificial Intelligence*, pages 203–208, Vancouver, BC, Canada.
- Candace L. Sidner. 1979. Toward a computational theory of definite anaphora comprehension in English. Technical Report AI-TR-537, MIT.
- Candace Sidner. 1983. What the speaker means: the recognition of speakers plans in discourse. *International Journal of Computers and Mathematics*, 9:71–82.
- Candace Sidner. 1985. Plan parsing for intended response recognition in discourse. *Computational Intelligence*, 1.
- Candace Sidner. 1992. A collaborative model of negotiation dialogues. *AAAI Workshop on Heterogeneous Agents*.
- Robert C. Stalnaker. 1978. Assertion. In Peter Cole, editor, *Syntax and Semantics, Volume 9: Pragmatics*, pages 315–332. Academic Press.
- Lucy Suchman. 1985. *Plans and Situated Actions*. Cambridge University Press.
- Deborah Tannen. 1989. *Talking voices : repetition, dialogue, and imagery in conversational discourse*. CUP.
- David R. Traum. 1991. The discourse reasoner in trains-90. TRAINS Technical Note 91-5, Computer Science Dept. University of Rochester.
- Moshe Y. Vardi. 1989. On the complexity of epistemic reasoning. In *Proc. 4th IEEE Symp. on Logic in Computer Science*, pages 243–252.
- Marilyn A. Walker and Steve Whittaker. 1990. Mixed initiative in dialogue: An investigation into discourse segmentation. In *Proc. 28th Annual Meeting of the ACL*, pages 70–79.
- Gregory Ward and Ellen Prince. 1991. On the topicalization of indefinite nps. *Journal of Pragmatics*, pages 338–351.
- Gregory Ward. 1985. *The syntax and semantics of preposing*. Ph.D. thesis, University of Pennsylvania.
- Gregory L. Ward. 1990. The discourse functions of vp preposing. *Language*, 66(4):742–763.
- Bonnie Lynn Webber and Aravind Joshi. 1982. Taking the initiative in natural language database interaction: Justifying why. Technical Report 1, Department of Computer and Information Science, University of Pennsylvania.
- Bonnie Lynn Webber. 1978. *A Formal Approach to Discourse Anaphora*. Ph.D. thesis, Harvard University. Garland Press.
- Bonnie Lynn Webber. 1986. Two steps closer to event reference. Technical Report MS-CIS-86-74, Linc Lab 42, Department of Computer and Information Science, University of Pennsylvania.

Steve Whittaker and Phil Stenton. 1988. Cues and control in expert client dialogues. In *Proc. 26th Annual Meeting of the ACL, Association of Computational Linguistics*, pages 123–130.

Steve Whittaker, Erik Geelhoed, and Elizabeth Robinson. 1992. Shared workspaces: How do they work and when are they useful? Technical report, HP Labs, Bristol, England.